Real-Life Math

Real-Life Math

Real-Life Math

Volume 1: A–L

K. Lee Lerner & Brenda Wilmoth Lerner, Editors

THOMSON GALE

Detroit • New York • San Francisco • San Diego • New Haven, Conn. • Waterville, Maine • London • Munich

Real-Life Math

Volume 2: M–Z

Field of Application Index
General Index

K. Lee Lerner & Brenda Wilmoth Lerner, Editors

THOMSON GALE

Detroit • New York • San Francisco • San Diego • New Haven, Conn. • Waterville, Maine • London • Munich

THOMSON GALE

Real-Life Math

K. Lee Lerner and Brenda Wilmoth Lerner, Editors

Project Editor
Kimberley A. McGrath

Editorial
Luann Brennan, Meggin M. Condino, Madeline Harris, Paul Lewon, Elizabeth Manar

Editorial Support Services
Andrea Lopeman

Indexing
Factiva, a Dow Jones & Reuters Company

Rights and Acquisitions
Margaret Abendroth, Timothy Sisler

Imaging and Multimedia
Lezlie Light, Denay Wilding

Product Design
Pamela Galbreath, Tracey Rowens

Composition
Evi Seoud, Mary Beth Trimper

Manufacturing
Wendy Blurton, Dorothy Maki

© 2006 Thomson Gale, a part of the Thomson Corporation.

Thomson and Star Logo are trademarks and Gale and UXL are registered trademarks used herein under license.

For more information, contact:
Thomson Gale
27500 Drake Rd.
Farmington Hills, MI 48331-3535
Or you can visit our Internet site at
http://www.gale.com

ALL RIGHTS RESERVED
No part of this work covered by the copyright hereon may be reproduced or used in any form or by any means—graphic, electronic, or mechanical, including photocopying, recording, taping, Web distribution, or information storage retrieval systems—without the written permission of the publisher.

For permission to use material from this product, submit your request via Web at http://www.gale-edit.com/permissions, or you may download our Permissions Request form and submit your request by fax or mail to:

Permissions
Thomson Gale
27500 Drake Rd.
Farmington Hills, MI 48331-3535
Permissions Hotline:
248-699-8006 or 800-877-4253, ext. 8006
Fax: 248-699-8074 or 800-762-4058

While every effort has been made to ensure the reliability of the information presented in this publication, Thomson Gale does not guarantee the accuracy of the data contained herein. Thomson Gale accepts no payment for listing; and inclusion in the publication of any organization, agency, institution, publication, service, or individual does not imply endorsement of the editors or publisher. Errors brought to the attention of the publisher and verified to the satisfaction of the publisher will be corrected in future editions.

LIBRARY OF CONGRESS CATALOGING-IN-PUBLICATION DATA

Real-life math / K. Lee Lerner and Brenda Wilmoth Lerner, editors.
p. cm.
Includes bibliographical references and index.
ISBN 0-7876-9422-3 (set : hardcover: alk. paper)—
ISBN 0-7876-9423-1 (v. 1)—ISBN 0-7876-9424-X (v. 2)
1. Mathematics—Encyclopedias.
I. Lerner, K. Lee. II. Lerner, Brenda Wilmoth.

QA5.R36 2006
510'.3—dc22 2005013141

This title is also available as an e-book, ISBN 1414404999 (e-book set).
ISBN: 0-7876-9422-3 (set); 0-7876-9423-1 (v1); 0-7876-9424-X (v2)
Contact your Gale sales representative for ordering information.

Printed in the United States of America
10 9 8 7 6 5 4 3 2 1

Table of Contents

Entries (With Areas of Discussion) . . . . . *vii*

Introduction . . . . . *xix*

List of Advisors and Contributors . . . . . *xxi*

Entries *1*

Volume 1: A–L

Addition 1
Algebra 9
Algorithms 26
Architectural Math 33
Area 45
Average. 51
Base 59
Business Math 62
Calculator Math 69
Calculus 80
Calendars 97
Cartography 100
Charts 107
Computers and Mathematics 114
Conversions 122
Coordinate Systems 131
Decimals 138
Demographics 141
Discrete Mathematics 144
Division 149
Domain and Range 156
Elliptic Functions 159
Estimation 161
Exponents 167
Factoring 180
Financial Calculations, Personal 184
Fractals 198
Fractions 203
Functions 210
Game Math 215
Game Theory 225
Geometry. 232
Graphing 248
Imaging 262
Information Theory 269
Inverse 278
Iteration 284
Linear Mathematics 287
Logarithms 294
Logic 300

Volume 2: M–Z

Matrices and Determinants . . . 303
Measurement . . . 307
Medical Mathematics . . . 314
Modeling . . . 328
Multiplication . . . 335
Music and Mathematics . . . 343
Nature and Numbers . . . 353
Negative Numbers . . . 356
Number Theory . . . 360
Odds . . . 365
Percentages . . . 372
Perimeter . . . 385
Perspective . . . 389
Photography Math . . . 398
Plots and Diagrams . . . 404
Powers . . . 416
Prime Numbers . . . 420
Probability . . . 423
Proportion . . . 430
Quadratic, Cubic, and Quartic Equations . . . 438
Ratio . . . 441
Rounding . . . 449
Rubric . . . 453
Sampling . . . 457
Scale . . . 465
Scientific Math . . . 473
Scientific Notation . . . 484
Sequences, Sets, and Series . . . 491
Sports Math . . . 495
Square and Cube Roots . . . 511
Statistics . . . 516
Subtraction . . . 529
Symmetry . . . 537
Tables . . . 543
Topology . . . 553
Trigonometry . . . 557
Vectors . . . 568
Volume . . . 575
Word Problems . . . 583
Zero-sum Games . . . 595
Glossary . . . 599
Field of Application Index . . . 605
General Index . . . 609

Entries (With Areas of Discussion)

Addition

Financial Addition 4
Geometric Progression 6
Poker, Probability, and Other Uses of Addition . . . 5
Sports and Fitness Addition 3
Using Addition to Predict and Entertain 6

Algebra

Art 21
Building Skyscrapers 19
Buying Light Bulbs 20
College Football 14
Crash Tests 18
Fingerprint Scanners 22
Flying an Airplane 16
Fundraising 19
Personal Finances 13
Population Dynamics 22
Private Space Travel 24
Skydiving 17
Teleportation. 23
UPC Barcodes 15

Algorithms

Archeology 27
Artificial Intelligence 32
Computer Programming 27
Credit Card Fraud Detection 27
Cryptology 28
Data Mining 28
Digital Animation and Digital Model Creation . . 28
DNA or Genetic Analysis 28
Encryption and Encryption Devices 28
The Genetic Code 30
Imaging 29
Internet Data Transmission 29
Linguistics, the Study of Language 31
Mapping 30
Market or Sales Analysis 30
Operational Algorithms 27
Security Devices 31
Sports Standings and Seedings 31
Tax Returns 31

Architectural Math

Architectural Concepts in Wheels 43
Architectural Symmetry in Buildings 38

Architecture 36
Astronomy 43
Ergonomics 41
Geometry, Basic Forms and Shapes of 40
Golden Rectangle and Golden Ratio 38
Grids, Use of 37
Jewelry 41
Measurement 35
Proportion 34
Ratio 33
Ratio and Proportion, Use of 38
Scale Drawing 34
Space, Use Of 37
Sports 38
Symmetry 34
Symmetry in City Planning 41
Technology 41
Textile and Fabrics 43

Area

Area of a Rectangle 45
Areas of Common Shapes 46
Areas of Solid Objects 46
Buying by Area 47
Car Radiators 48
Cloud and Ice Area and Global Warming 47
Drug Dosing 46
Filtering 47
Solar Panels 49
Surveying 48
Units of Area 45

Average

Arithmetic Mean 51
The "Average" Family 55
Average Lifespan 57
Averaging for Accuracy 55
Batting Averages 53
Evolution in Action 57
Geometric Mean 52
Grades 54
How Many Galaxies? 55
Insurance 57
Mean 52
Median 52
Space Shuttle Safety 56
Student Loan Consolidation 56
Weighted Averages in Business 54
Weighted Averages in Grading 54

Base

Base 2 and Computers 60

Business Math

Accounting 63
Budgets 63
Earnings 66
Interest 67
Payroll 65
Profits 66

Calculator Math

Bridge Construction 76
Combinatorics 77
Compound Interest 74
Financial Transactions 73
Measurement Calculations 75
Nautical Navigation 73
Random Number Generator 75
Supercomputers 78
Understanding Weather 77

Calculus

Applications of Derivatives 86
Derivative 81
Functions 81
Fundamental Theorem of Calculus 85
Integral 83
Integrals, Applications 91
Maxima and Minima 85

Calendars

Gregorian Calendar 99
Islamic and Chinese Calendars 99
Leap Year 99

Cartography

Coordinate Systems 103
GIS-Based Site Selection 105
GPS Navigation 105
Map Projection 100
Natural Resources Evaluation and Protection . . . 105
Scale 100
Topographic Maps 104

Charts

Bar Charts . . . 109
Basic Charts . . . 107
Choosing the Right Type of Chart For the Data . . . 112
Clustered Column Charts . . . 110
Column and Bar Charts . . . 109
Line Charts . . . 107
Pie Charts . . . 110
Stacked Column Charts . . . 110
Using the Computer to Create Charts . . . 112
X-Y Scatter Graphs . . . 109

Computers and Mathematics

Algorithms . . . 115
Binary System . . . 114
Bits . . . 116
Bytes . . . 116
Compression. . . . 118
Data Transmission . . . 119
Encryption . . . 120
IP Address . . . 117
Pixels, Screen Size, and Resolution . . . 117
Subnet Mask. . . . 118
Text Code . . . 116

Conversions

Absolute Systems . . . 127
Arbitrary Systems . . . 128
Cooking or Baking Temperatures . . . 127
Derived Units . . . 124
English System . . . 123
International System of Units (SI) . . . 123
Metric Units . . . 123
Units Based On Physical or "Natural" Phenomena . . . 124
Weather Forecasting . . . 126

Coordinate Systems

3-D Systems On Ordinance Survey Maps . . . 136
Cartesian Coordinate Plane . . . 132
Changing Between Coordinate Systems . . . 132
Choosing the Best Coordinate System . . . 132
Commercial Aviation . . . 135
Coordinate Systems Used in Board Games . . . 134
Coordinate Systems Used for Computer Animation . . . 134
Dimensions of a Coordinate System . . . 131
Longitude and John Harrison . . . 135
Modern Navigation and GPS . . . 135
Paper Maps of the World . . . 134
Polar Coordinates . . . 133
Radar Systems and Polar Coordinates . . . 136
Vectors . . . 132

Decimals

Grade Point Average Calculations . . . 139
Measurement Systems . . . 139
Science . . . 139

Demographics

Census . . . 142
Election Analysis . . . 141
Geographic Information System Technology . . . 143

Discrete Mathematics

Algorithms . . . 145
Boolean Algebra . . . 145
Combinatorial Chemistry . . . 147
Combinatorics . . . 145
Computer Design . . . 146
Counting Jaguars Using Probability Theory . . . 147
Cryptography . . . 146
Finding New Drugs with Graph Theory . . . 147
Graphs . . . 146
Logic, Sets, and Functions . . . 144
Looking Inside the Body With Matrices . . . 147
Matrix Algebra . . . 146
Number Theory . . . 145
Probability Theory. . . . 145
Searching the Web . . . 146
Shopping Online and Prime Numbers . . . 147

Division

Averages . . . 152
Division and Comparison . . . 151
Division and Distribution . . . 150
Division, Other Uses . . . 153
Practical Uses of Division For Students . . . 153

Domain and Range

Astronomers . . . 157
Calculating Odds and Outcomes . . . 157

Computer Control and Coordination 157
Computer Science 158
Engineering 157
Graphs, Charts, Maps 158
Physics 157

Elliptic Functions

The Age of the Universe 160
Conformal Maps 159
E-Money 160

Estimation

Buying a Used Car 162
Carbon Dating 165
Digital Imaging 164
Gumball Contest 163
Hubble Space Telescope 165
Population Sampling 164
Software Development 166

Exponents

Bases and Exponents 167
Body Proportions and Growth
(Why Elephants Don't Have
Skinny Legs) 179
Credit Card Meltdown 178
Expanding Universe 178
Exponential Functions 168
Exponential Growth 171
Exponents and Evolution 174
Integer Exponents 167
Interest and Inflation 177
Non-Integer Exponents 168
Radioactive Dating 177
Radioactive Decay 175
Rotting Leftovers 173
Scientific Notation 171

Factoring

Codes and Code Breaking 182
Distribution 182
Geometry and Approximation of Size 182
Identification of Patterns
and Behaviors 181
Reducing Equations 181
Skill Transfer 182

Financial Calculations, Personal

Balancing a Checkbook 189
Budgets 188
Buying Music 184
Calculating a Tip 194
Car Purchasing and Payments 187
Choosing a Wireless Plan 187
Credit Cards 185
Currency Exchange 195
Investing 190
Retirement Investing 192
Social Security System 190
Understanding Income Taxes 189

Fractals

Astronomy 202
Building Fractals 199
Cell Phone and Radio Antenna 202
Computer Science 202
Fractals and Nature 200
Modeling Hurricanes and Tornadoes 201
Nonliving Systems 201
Similarity 199

Fractions

Algebra 205
Cooking and Baking 206
Fractions and Decimals 204
Fractions and Percentages 204
Fractions and Voting 208
Music 206
Overtime Pay 208
Radioactive Waste 206
Rules For Handling Fractions 204
Simple Probabilities 207
Tools and Construction 208
Types of Fractions 203
What Is a Fraction? 203

Functions

Body Mass Index 214
Finite-Element Models 212
Functions, Described 210
Functions and Relations 210
Guilloché Patterns 211
Making Airplanes Fly 211

The Million-Dollar Hypothesis 212
Nuclear Waste 213
Synths and Drums 213

Game Math

Basic Board Games. 220
Card Games 218
Magic Squares 221
Math Puzzles 223
Other Casino Games 219

Game Theory

Artificial Intelligence 230
Decision Theory 228
eBay and the Online Auction World 230
Economics 229
Economics and Game Theory 228
Evolution and Animal Behavior 229
General Equilibrium 229
Infectious Disease Therapy 230
Nash Equilibrium 229

Geometry

Architecture 237
Fireworks 241
Fourth Dimension 245
Global Positioning 239
Honeycombs 239
Manipulating Sound 241
Pothole Covers 236
Robotic Surgery. 245
Rubik's Cube 243
Shooting an Arrow 244
Solar Systems 242
Stealth Technology 244

Graphing

Aerodynamics and Hydrodynamics 259
Area Graphs 252
Bar Graphs 249
Biomedical Research 258
Bubble Graphs 257
Computer Network Design 259
Finding Oil 258
Gantt Graphs 254
Global Warming 257
GPS Surveying 258
Line Graphs 251
Physical Fitness 259
Picture Graphs 254
Pie Graphs 252
Radar Graphs 253
X-Y Graphs 254

Imaging

Altering Images 263
Analyzing Images 263
Art 267
Compression 264
Creating Images 263
Dance 266
Forensic Digital Imaging 266
Meat and Potatoes 266
Medical Imaging 264
Optics 264
Recognizing Faces: a Controversial Biometrics Application 264
Steganography and Digital Watermarks 266

Information Theory

Communications 273
Error Correction 275
Information and Meaning 273
Information Theory in Biology and Genetics . . . 274
Quantum Computing 276
Unequally Likely Messages 271

Inverse

Anti-Sound 282
The Brain and the Inverted Image On the Eye 281
Cryptography 280
Definition of an Inverse 278
Fluid Mechanics and Nonlinear Design 281
Inverse Functions 279
The Multiplicative Inverse 278
Negatives Used in Photography 281
Operations Where the Inverse Does Not Exist 279
Operations With More Than One Inverse 279
Stealth Submarine Communications 282
Stereo 282

Iteration

Iteration and Business 285
Iteration and Computers 286
Iteration and Creativity 285
Iteration and Sports 284

Linear Mathematics

Earthquake Prediction 289
Linear Programming 291
Linear Reproduction of Music 292
Recovering Human Motion From Video 290
Virtual Tennis 291

Logarithms

Algebra of Powers of Logarithms 296
Computer Intensive Applications 297
Cryptography and Group Theory 299
Designing Radioactive Shielding
for Equipment in Space 299
Developing Optical Equipment. 298
Estimating the Age of Organic Matter
Using Carbon Dating 298
Log Tables 296
Logarithms to Other Bases Than 10 296
The Power of Mathematical Notation 295
Powers and Logs of Base 10 295
Powers and Their Relation
to Logarithms 296
Supersonic and Hypersonic Flight 299
Use in Medical Equipment 298
Using a Logarithmic Scale to Measure
Sound Intensity 297

Logic

Boolean Logic 300
Fuzzy Logic 300
Proposition and Conclusion 300
Reasoning 300

Matrices and Determinants

Designing Cars 305
Digital Images 304
Flying the Space Shuttle 305
Population Biology 305

Measurement

Accuracy in Measurement 309
Archaeology 310
Architecture 310
Blood Pressure 310
Chemistry 310
Computers 310
The Definition of a Second 310
Dimensions 308
Doctors and Medicine. 310
Engineering 309
Evaluating Errors in Measurement
and Quality Control 309
Gravity 313
How Astronomers and NASA Measure
Distances in Space 312
Measuring Distance 308
Measuring Mass 313
Measuring the Speed of Gravity 313
Measuring Speed, Space Travel, and Racing . . . 310
Measuring Time 310
Navigation 310
Nuclear Power Plants 310
Space Travel and Timekeeping 312
Speed of Light 312

Medical Mathematics

Calculation of Body Mass Index (BMI) 319
Clinical Trials 323
Genetic Risk Factors: the Inheritance of Disease . . 321
Rate of Bacterial Growth 326
Standard Deviation and Variance for Use
in Height and Weight Charts 319
Value of Diagnostic Tests 318

Modeling

Ecological Modeling 330
Military Modeling 331

Multiplication

Sports Multiplication: Calculating
a Baseball ERA 338
Calculating Exponential Growth Rates 338
Calculating Miles Per Gallon 341
Electronic Timing 339
Exponents and Growth Rates 337
Investment Calculations 337

Measurement Systems 339
Multiplication in International Travel 339
Other Uses of Multiplication 340
Rate of Pay 339
Savings 341
SPAM and Email Communications 341

Music and Mathematics

Acoustic Design 348
Compressing Music 349
Computer-Generated Music 349
Digital Music. 348
Discordance of the Spheres 346
Electronic Instruments 347
Error Correction 349
Frequency of Concert A 351
Mathematical Analysis of Sound 347
Math-Rock 351
Medieval Monks 345
Pythagoras and Strings 343
Quantification of Music 345
Using Randomness 349
Well-Tempered Tones 346

Nature and Numbers

Fibonacci Numbers and the Golden Ratio 353
Mathematical Modeling of Nature 354
Specify Application Using Alphabetizable Title . . 355
Using Fractals to Represent Nature 355

Negative Numbers

Accounting Practice 357
Buildings 359
Flood Control 358
The Mathematics of Bookkeeping 357
Sports 358
Temperature Measurement 357

Number Theory

Cryptography 362
Error Correcting Codes 363

Odds

Odds in Everyday Life 367
Odds in State Lotteries 368
Odds, Other Applications 369
Sports and Entertainment Odds 366

Percentages

Calculating a Tip 375
Compound Interest 376
Definitions and Basic Applications 372
Examples of Common Percentage
Applications 374
Finding the Base Rate 374
Finding the Original Amount 375
Finding the Rate of Increase or Decrease 375
Finding the Rate Percent 374
Important Percentage Applications 374
Percentage Change: Increase or Decrease 375
Public Opinion Polls 379
Ratios, Proportions, and Percentages 373
Rebate Period and Cost 378
Rebates 377
Retail Sales: Price Discounts and Markups
and Sales Tax 376
Sales Tax Calculation: In-Store Discount
Versus Mail-In Rebate 377
Sales Tax Calculations 377
SAT Scores or Other Academic Testing 383
Sports Math 379
Tournaments and Championships 382
Understanding Percentages in the Media 378
Using Percentages to Make Comparisons 379

Perimeter

Bodies of Water 386
Landscaping 386
Military 387
Planetary Exploration 388
Robotic Perimeter Detection Systems 388
Security Systems 386
Sporting Events 386

Perspective

Animation 392
Art 391
Computer Graphics 395
Film 393
Illustration 392
Interior Design 394
Landscaping 395

Photography Math

The Camera . . . 398
Depth of Field . . . 400
Digital Image Processing . . . 403
Digital Photography . . . 401
Film Speed . . . 398
Lens Aperture . . . 400
Lens Focal Length . . . 399
Photomicrography . . . 403
Reciprocity . . . 401
Shutter Speed . . . 399
Sports and Wildlife Photography . . . 402

Plots and Diagrams

Area Chart . . . 406
Bar Graphs . . . 406
Body Diagram . . . 414
Box Plot . . . 405
Circuit Diagram. . . . 414
Diagrams . . . 404
Fishbone Diagram . . . 406
Flow Chart . . . 411
Gantt Charts . . . 413
Line Graph . . . 408
Maps . . . 413
Organization Charts . . . 413
Other Diagrams . . . 414
Pie Graph. . . . 406
Polar Chart . . . 406
Properties of Graphs . . . 404
Scatter Graph . . . 405
Stem and Leaf Plots . . . 405
Street Signs . . . 414
Three-Dimensional Graph . . . 407
Tree Diagram . . . 412
Triangular Graph . . . 407
Weather Maps . . . 414

Powers

Acids, Bases, and pH Level . . . 418
Areas of Polygons and Volumes of Solid Figures . . . 417
Astronomy and Brightness of Stars . . . 418
Computer Science and Binary Logic . . . 417
Earthquakes and the Richter Scale . . . 417
The Powers of Nanotechnology . . . 418

Prime Numbers

Biological Applications of Prime Numbers . . . 421

Probability

Gambling and Probability Myths . . . 425
Probability in Business and Industry . . . 427
Probability, Other Uses . . . 428
Probability in Sports and Entertainment . . . 426
Security . . . 424

Proportion

Architecture . . . 432
Art, Sculpture, and Design . . . 432
Chemistry . . . 435
Diets . . . 436
Direct Proportion . . . 431
Engineering Design . . . 435
Ergonomics . . . 434
Inverse Proportion . . . 431
Maps . . . 434
Medicine . . . 434
Musical Instruments . . . 435
Proportion in Nature . . . 436
Solving Ratios With Cross Products . . . 430
Stock Market . . . 436

Quadratic, Cubic, and Quartic Equations

Acceleration . . . 439
Area and Volume . . . 439
Car Tires . . . 439
Guiding Weapons . . . 440
Hospital Size . . . 440
Just in Time Manufacturing . . . 440

Ratio

Age of Earth . . . 446
Automobile Performance . . . 445
Cleaning Water . . . 446
Cooking . . . 446
Cost of Gas . . . 443
Determination of the Origination of the Moon . . . 447
Genetic Traits . . . 443
Healthy Living . . . 446
Length of a Trip . . . 443
Music . . . 445

Optimizing Livestock Production 447
Sports 445
Stem Cell Research. 446
Student-Teacher Ratio 445

Rounding

Accounting 451
Bulk Purchases 450
Decimals 450
Energy Consumption 451
Length and Weight 450
Lunar Cycles 451
Mileage 452
Pi 450
Population 451
Precision 452
Time 452
Weight Determination 451
Whole Numbers. 449

Rubric

Analytic Rubrics and Holistic Rubrics 455
General Rubrics and Task-Specific Rubrics 455
Scoring Rubrics 453

Sampling

Agriculture 459
Archeology 463
Astronomy 462
Demographic Surveys 462
Drug Manufacturing 460
Environmental Studies 462
Market Assessment 463
Marketing 463
Non-Probability Sampling 458
Plant Analysis 460
Probability Sampling 457
Scientific Research 460
Soil Sampling 460
Weather Forecasts 461

Scale

Architecture 468
Atmospheric Pressure Using Barometer 469
The Calendar 469
Expanse of Scale From the Sub-Atomic to the Universe 471
Interval Scale 466
Linear Scale 465
Logarithmic Scale 465
Map Scale 467
Measuring Wind Strength 469
The Metric System of Measurement 472
Music 471
Nominal Scale 467
Ordinal Scale 467
Ratio Scale 466
The Richter Scale 470
Sampling 472
Technology and Imaging 469
Toys 471
Weighing Scale 468

Scientific Math

Aviation and Flights 478
Bridging Chasms 478
Discrete Math 474
Earthquakes and Logarithms 482
Equations and Graphs 476
Estimating Data Used for Assessing Weather . . . 477
Functions and Measurements 473
Genetics 483
Logarithms 475
Matrices and Arrays 475
Medical Imaging 480
Rocket Launch 480
Ships 482
Simple Carpentry 479
Trigonometry and the Pythagorean Theorem . . . 474
Weather Prediction 476
Wind Chill in Cold Weather 476

Scientific Notation

Absolute Dating 489
Chemistry 486
Computer Science 487
Cosmology 487
Earth Science 489
Electrical Circuits 486
Electronics 489
Engineering 487
Environmental Science 488
Forensic Science 488
Geologic Time Scale and Geology 488

Light Years, the Speed of Light, and Astronomy . . . 486
Medicine . . . 488
Nanotechnology . . . 490
Proteins and Biology . . . 490

Sequences, Sets, and Series

Genetics . . . 493
Operating On Sets . . . 492
Ordering Things . . . 493
Sequences . . . 491
Series . . . 492
Sets . . . 491
Using Sequences . . . 493

Sports Math

Baseball . . . 498
Basketball . . . 499
Cycling—Gear Ratios and How They Work . . . 505
Football—How Far was the Pass Thrown? . . . 507
Football Tactics—Math as a Decision-Making Tool . . . 501
Golf Technology . . . 506
Math and the Science of Sport . . . 504
Math and Sports Wagering . . . 508
Math to Understand Sports Performance . . . 497
Mathematics and the Judging of Sports . . . 504
Money in Sport—Capology 101 . . . 507
North American Football . . . 499
Pascal's Triangle and Predicting a Coin Toss . . . 500
Predicting the Future: Calling the Coin Toss . . . 500
Ratings Percentage Index (RPI) . . . 503
Rules Math . . . 496
Soccer—Free Kicks and the Trajectory of the Ball . . . 506
Understanding the Sports Media Expert . . . 502

Square and Cube Roots

Architecture . . . 513
Global Economics . . . 515
Hiopasus's Fatal Discovery . . . 513
Names and Conventions . . . 512
Navigation . . . 514
Pythagorean Theorem . . . 513
Sports . . . 514
Stock Markets . . . 515

Statistics

Analysis of Variance . . . 522
Average Values . . . 519
Confidence Intervals . . . 522
Correlation and Curve Fitting . . . 521
Cumulative Frequencies and Quantiles . . . 521
Geostatistics . . . 525
Measures of Dispersion . . . 520
Minimum, Maximum, and Range . . . 518
Populations and Samples . . . 516
Probability . . . 517
Public Opinion Polls . . . 527
Quality Assurance . . . 526
Statistical Hypothesis Testing . . . 522
Using Statistics to Deceive . . . 523

Subtraction

Subtraction in Entertainment and Recreation . . . 533
Subtraction in Financial Calculations . . . 531
Subtraction in Politics and Industry . . . 535
Tax Deductions . . . 532

Symmetry

Architecture . . . 541
Exploring Symmetries . . . 539
Fractal Symmetries . . . 541
Imperfect Symmetries . . . 542
Symmetries in Nature . . . 542

Tables

Converting Measurements . . . 545
Daily Use . . . 549
Educational Tables . . . 545
Finance . . . 546
Health . . . 548
Math Skills . . . 544
Travel . . . 549

Topology

Computer Networking . . . 555
I.Q. Tests . . . 555
Möbius Strip . . . 555
Visual Analysis . . . 554
Visual Representation . . . 555

Trigonometry

Chemical Analysis 566
Computer Graphics 566
Law of Sines 561
Measuring Angles 557
Navigation 562
Pythagorean Theorem. 559
Surveying, Geodesy, and Mapping 564
Trigonometric Functions 560
Types of Triangles 558
Vectors, Forces, and Velocities 563

Vectors

3-D Computer Graphics 572
Drag Racing 572
Land Mine Detection 572
The Magnitude of a Vector 569
Sports Injuries 573
Three-Dimensional Vectors 569
Two-Dimensional Vectors 568
Vector Algebra 570
Vectors in Linear Algebra 571

Volume

Biometric Measurements. 581
Building and Architecture 579
Compression Ratios in Engines 579
Glowing Bubbles: Sonoluminescence 579
Medical Applications 578
Misleading Graphics 581
Pricing 577
Runoff 582
Sea Level Changes 580
Swimming Pool Maintenance 581
Units of Volume 575
Volume of a Box 575
Volumes of Common Solids 575
Why Thermometers Work 580

Word Problems

Accounts and VAT 592
Archaeology 585
Architecture 590
Average Height? 593
Banks, Interest Rates, and Introductory Rates . . . 591
Bearings and Directions of Travel 592
Comparisons 586
Computer Programming 584
Cooking Instructions 591
Creative Design 584
Cryptography 585
Decorating 594
Disease Control 591
Ecology 587
Efficient Packing and Organization 590
Engineering 585
Exchange Rates 586
Finance 591
Geology 591
Global Warming 594
Graph Theory 588
Hypothesis Testing 585
Insurance 584
Linear Programming 588
Lotteries and Gambling 591
Measuring Height of a Well 594
Medicine and Cures 585
MMR Immunization and Autism 594
Navigation 587
Opinion Polls 593
Percentages 586
Phone Companies 586
Postman 589
Proportion and Inverse Proportion 586
Quality Control 592
Ranking Test Scores 589
Recipes 591
ROTA and Timetables 589
Searching in an Index 590
Seeding in Tournaments 590
Shortest Links to Establish Electricity
to a Whole Town 589
Software Design 584
Stock Keeping 592
Store Assistants 592
Surveying 592
Teachers 584
Throwing a Ball 593
Translation 587
Travel and Racing 586
Traveling Salesperson 589
Weather 593

Zero-Sum Games

Currency, Futures, and Stock Markets 597
Experimental Gaming 597
Gambling 596
War 597

Introduction

Real-Life Math takes an international perspective in exploring the role of mathematics in everyday life and is intended for high school age readers. As *Real-Life Math* (*RLM*) is intended for a younger and less mathematically experienced audience, the authors and editors faced unique challenges in selecting and preparing entries.

The articles in the book are meant to be understandable by anyone with a curiosity about mathematical topics. *Real-Life Math* is intended to serve all students of math such that an 8th- or 9th-grade student just beginning their study of higher maths can at least partially comprehend and appreciate the value of courses to be taken in future years. Accordingly, articles were constructed to contain material that might serve all students. For example, the article, "Calculus" is intended to be able to serve students taking calculus, students finished with prerequisites and about to undertake their study of calculus, and students in basic math or algebra who might have an interest in the practical utility of a far-off study of calculus. Readers should anticipate that they might be able to read and reread articles several times over the course of their studies in maths. *Real-Life Math* challenges students on multiple levels and is designed to facilitate critical thinking and reading-in-context skills. The beginning student is not expected to understand more mathematically complex text dealing, for example, with the techniques for calculus, and so should be content to skim through these sections as they read about the practical applications. As students progress through math studies, they will naturally appreciate greater portions of more advanced sections designed to serve more advanced students.

To be of maximum utility to students and teachers, most of the 80 topics found herein—arranged alphabetically by theory or principle—were predesigned to correspond to commonly studied fundamental mathematical concepts as stated in high school level curriculum objectives. However, as high school level maths generally teach concepts designed to develop skills toward higher maths of greater utility, this format sometimes presented a challenge with regard to articulating understandable or direct practical applications for fundamental skills without introducing additional concepts to be studied in more advanced math classes. It was sometimes difficult to isolate practical applications for fundamental concepts because it often required more complex mathematical concepts to most accurately convey the true relationship of mathematics to our advancing technology. Both the authors and editors of the project made exceptional efforts to smoothly and seamlessly incorporate the concepts necessary (and at an accessible level) within the text.

Although the authors of *Real-Life Math* include math teachers and professors, the bulk of the writers are

practicing engineers and scientists who use math on a daily basis. However, *RLM* is not intended to be a book about real-life applications as used by mathematicians and scientists but rather, wherever possible, to illustrate and discuss applications within the experience—and that are understandable and interesting—to younger readers.

RLM is intended to maximize readability and accessibility by minimizing the use of equations, example problems, proofs, etc. Accordingly, *RLM* is not a math textbook, nor is it designed to fully explain the mathematics involved in each concept. Rather, *RLM* is intended to compliment the mathematics curriculum by serving a general reader for maths by remaining focused on fundamental math concepts as opposed to the history of math, biographies of mathematicians, or simply interesting applications. To be sure, there are inherent difficulties in presenting mathematical concepts without the use of mathematical notation, but the authors and editors of *RLM* sought to use descriptions and concepts instead of mathematical notation, problems, and proofs whenever possible.

To the extent that *RLM* meets these challenges it becomes a valuable resource to students and teachers of mathematics.

The editors modestly hope that *Real-Life Math* serves to help students appreciate the scope of the importance and influence of math on everyday life. *RLM* will achieve its highest purposes if it intrigues and inspires students to continue their studies in maths and so advance their understanding of the both the utility and elegance of mathematics.

"[The universe] cannot be read until we have learnt the language and become familiar with the characters in which it is written. It is written in mathematical language, and the letters are triangles, circles, and other geometrical figures, without which means it is humanly impossible to comprehend a single word." Galilei, Galileo (1564–1642)

K. Lee Lerner and Brenda Wilmoth Lerner, Editors

List of Advisors and Contributors

In compiling this edition, we have been fortunate in being able to rely upon the expertise and contributions of the following scholars who served as contributing advisors or authors for *Real-Life Math*, and to them we would like to express our sincere appreciation for their efforts:

William Arthur Atkins

Mr. Atkins holds a BS in physics and mathematics as well as an MBA. He lives at writes in Perkin, Illinois.

Juli M. Berwald, PhD

In addition to her graduate degree in ocean sciences, Dr. Berwald holds a BA in mathematics from Amherst College, Amherst, Massachusetts. She currently lives and writes in Chicago, Illinois.

Bennett Brooks

Mr. Brooks is a PhD graduate student in mathematics. He holds a BS in mathematics, with departmental honors, from University of Redlands, Redlands, California, and currently works as a writer based in Beaumont, California.

Rory Clarke, PhD

Dr. Clark is a British physicist conducting research in the area of high-energy physics at the University of Bucharest, Romania. He holds a PhD in high energy particle physics from the University of Birmingham, an MSc in theoretical physics from Imperial College, and a BSc degree in physics from the University of London.

Raymond C. Cole

Mr. Cole is an investment banking financial analyst who lives in New York. He holds an MBA from the Baruch Zicklin School of Business and a BS in business administration from Fordham University.

Bryan Thomas Davies

Mr. Davies holds a Bachelor of Laws (LLB) from the University of Western Ontario and has served as a criminal prosecutor in the Ontario Ministry of the Attorney General. In addition to his legal experience, Mr. Davies is a nationally certified basketball coach.

John F. Engle

Mr. Engle is a medical student at Tulane University Medical School in New Orleans, Louisiana.

William J. Engle
Mr. Engle is a retired petroleum engineer who lives in Slidell, Louisiana.

Paul Fellows
Dr. Fellows is a physicist and mathematician who lives in London, England.

Renata A. Ficek
Ms. Ficek is a graduate mathematics student at the University of Queensland, Australia.

Larry Gilman, PhD
Dr. Gilman holds a PhD in electrical engineering from Dartmouth College and an MA in English literature from Northwestern University. He lives in Sharon, Vermont.

Amit Gupta
Mr. Gupta holds an MS in information systems and is managing director of Agarwal Management Consultants P. Ltd., in Ahmedabad, India.

William C. Haneberg, PhD
Dr. Haneberg is a professional geologist and writer based in Seattle, Washington.

Bryan D. Hoyle, PhD
Dr. Hoyle is a microbiologist and science writer who lives in Halifax, Nova Scotia, Canada.

Kenneth T. LaPensee, PhD
In addition to professional research in epidemiology, Dr. LaPensee directs Skylands Healthcare Consulting located in Hampton, New Jersey.

Holly F. McBain
Ms. McBain is a science and math writer who lives near New Braunfels, Texas.

Mark H. Phillips, PhD
Dr. Phillips serves as an assistant professor of management at Abilene Christian University, located in Abilene, Texas.

Nephele Tempest
Ms. Tempest is a writer based in Los Angeles, California.

David Tulloch
Mr. Tulloch holds a BSc in physics and an MS in the history of science. In addition to research and writing he serves as a radio broadcaster in Ngaio, Wellington, New Zealand.

James A. Yates
Mr. Yates holds a MMath degree from Oxford University and is a teacher of maths in Skegnes, England.

ACKNOWLEDGMENTS

The editors would like to extend special thanks to Connie Clyde for her assistance in copyediting. The editors also wish to especially acknowledge Dr. Larry Gilman for his articles on calculus and exponents as well as his skilled corrections of the entire text. The editors are profoundly grateful to their assistant editors and proofreaders, including Lynn Nettles and Bill Engle, who read and corrected articles under the additional pressures created by evacuations mandated by Hurricane Katrina. The final editing of this book was interrupted as Katrina damaged the Gulf Coast homes and offices of several authors, assistant editors, and the editors of *RLM* just as the book was being prepared for press. Quite literally, many pages were read and corrected by light produced by emergency generators—and in some cases, pages were corrected from evacuation shelters. The editors are forever grateful for the patience and kind assistance of many fellow scholars and colleagues during this time.

The editors gratefully acknowledge the assistance of many at Thompson Gale for their help in preparing *Real-Life Math*. The editors wish to specifically thank Ms. Meggin Condino for her help and keen insights while launching this project. The deepest thanks are also offered to Gale Senior Editor Kim McGrath for her tireless, skilled, good-natured, and intelligent guidance.

Addition

Overview

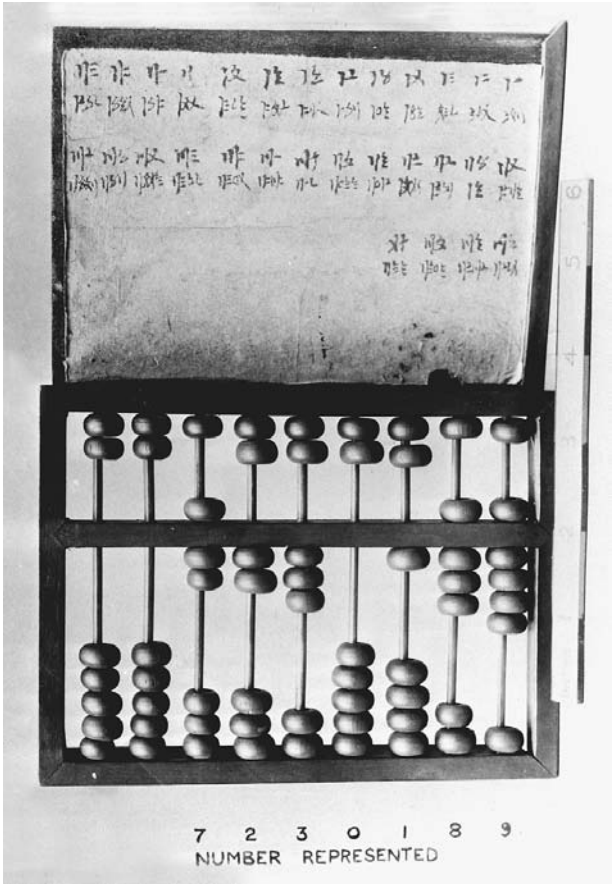
Addition is the process of combining two or more numbers to create a new value, and is generally considered the simplest form of mathematics. Despite its simplicity, the ability to perform basic addition is the foundation of most advanced mathematics, and simple addition, repeated millions of times per second, actually underlies much of the processing performed within the most advanced electronic computers on earth. Despite its elementary nature, the process of adding numbers together remains one of the most useful mathematical operations available, as well as perhaps the most common type of calculation performed on a daily basis by most adults.

Fundamental Mathematical Concepts and Terms

An addition equation requires only two terms to describe its component parts. When asked to name the simplest equation possible, most adults would respond with $1 + 1 = 2$, probably the first math operation they learned. In this simple equation, the two 1s are termed addends, while the result of this or any other addition equation is known as the sum, in this case the value 2. Because this final value is called a sum, it is also correct, though less common, to describe the process of adding as summing, as in the expression, "Sum the five daily values to find the total attendance for the week." While the addition sign is properly called a plus sign, one does not ever refer to the process of addition as "plus-ing" two values.

A Brief History of Discovery and Development

Because the basic process of addition is so simple, its exact origins are impossible to identify. Near the beginning of recorded history, a variety of endeavors including commerce, warfare, and agriculture required the ability to add numbers; for some lines of work, addition was such a routine operation that specific tools became necessary in order to streamline the process. The most basic counting tools consisted of a small bag of stones or other small objects that could be used to tally an inventory of goods. In the case of shipping, a merchant counting sacks of grain as they were loaded onto his ship would move one small stone aside for each sack loaded, providing both a running total and a simple method to double-check the final count. Upon arrival, this same collection of stones would serve as the ship's manifest, allowing a



The Chinese abacus was one of the earliest tools for everyday addition. CORBIS-BETTMANN. REPRODUCED BY PERMISSION.

running count of the shipment as it was unloaded. In the case of warfare, a general might number his horses using this same method of having each object represented by a stone, a small seashell, or some other token. The key principle in this type of system was a one-to-one relationship between the items being counted and the smaller symbolic items used to maintain the tally.

Over time, these sets of counting stones gradually evolved into large counting tables, known as abaci, or in the singular form, an abacus. These tables often featured grooves or other placement aids designed to insure accuracy in the calculations being made, and tallies were made by placing markers in the proper locations to symbolize ones, tens, and hundreds. The counting tables developed in numerous cultures, and ancient examples survive from Japan, Greece, China, and the Roman Empire. Once these tables came into wide use, a natural evolution, much like that seen in modern computer systems, occurred, with the bulky, fixed tables gradually morphing into smaller, more portable devices. These smaller versions were actually the earliest precursors of today's personal calculator.

The earliest known example of what we today recognize as the hand-held abacus was invented in China approximately 5,000 years ago. Consisting of wood and moveable beads, this counting tool did not actually perform calculations, but instead assisted its human operator by keeping a running total of items added. The Chinese abacus was recognized as an exceptionally useful tool, and progressively spread throughout the world. Modern examples of the abacus are little changed from these ancient models, and are still used in some parts of the world, where an expert user can often solve lengthy addition problems as quickly as someone using an electronic calculator.

As technology advanced, users sought ways to add more quickly and more accurately. In 1642, a French mathematician Blaise Pascal (1623–1662) invented the first mechanical adding machine. This device, a complex contraption operated by gears and wheels, allowed the user to type in his equation using a series of keys, with the results of the calculation displayed in a row of windows. Pascal's invention was revolutionary, specifically because it could carry digits from one column to another. Mechanical calculators, the distant descendents of Pascal's design, remained popular well into the twentieth century; more advanced electrically operated versions were used well into the 1960s and 1970s, when they were replaced by electronic models and spreadsheet software.

In a strange case of history repeating itself, the introduction of the first high-priced electronic calculators in the 1970s was coincidentally accompanied by television commercials offering training in a seemingly revolutionary method of adding called Chisenbop. Chisenbop allowed one to use only his fingers to add long columns of numbers very quickly, and television shows of that era featured young experts out-performing calculator-wielding adults. Chisenbop uses a variety of finger combinations to represent different values, with the right hand tallying values from zero to nine, and the left hand handling values from ten and up. The rapid drop in calculator prices during this era, as well as the potential stigma associated with counting on one's fingers, probably led to the method's demise. Despite its seemingly revolutionary nature, this counting scheme is actually quite old, and may in fact predate the abacus, which functions in a similar manner by allowing the operator to tally values as they are added. Multiple online tutorials today teach the technique, which has gradually faded back into obscurity.

While the complex calculations performed by today's sophisticated computers might appear to lie far beyond anything achieved by Pascal's original adding machine, the remnants of Pascal's simple additions can still be

found deep inside every microprocessor (as well as in a simple programming language which bears his name in honor of his pioneering work). Modern computers offer user-friendly graphic interfaces and require little or no math or programming knowledge on the part of the average user. But at the lowest functional level, even a cutting edge processor relies on simple operations performed in its arithmetic logic unit, or ALU. When this basic processing unit receives an instruction, that instruction has typically been broken down into a series of simple processes which are then completed one at a time. Ironically, though the ALU is the mathematical heart of a modern computer, a typical ALU performs only four functions, the same add, subtract, multiply, and divide found on the earliest electronic calculators of the 1970s. By performing these simple operations millions of times each second, and leveraging this power through modern operating systems and applications software, even a process as simple as addition can produce startling results.

Real-life Applications

SPORTS AND FITNESS ADDITION

Many aspects of popular sports require the use of addition. For example, some of the best-known records tracked in most sports are found by simply adding one success to another. Records for the most homeruns, the most 3-point shots made, the most touchdown passes completed, and the most major golf tournaments won in a career are nothing more than the result of lengthy addition problems stretched out over an entire career. On the business side of sports are other addition applications, including such routine tasks as calculating the number of fans at a ballgame or the number of hotdogs sold, both of which are found by simply adding one more person or sausage to the running total.

Many sports competitions are scored on the basis of elapsed time, which is found by simply adding fractions of a second to a total until the event ends, at which time the smallest total is determined to be the winning score. In the case of motor sports, racers compete for the chance to start the actual race near the front of the field, and these qualifying attempts are often separated by mere hundredths or even thousandths of a second. Track events such as the decathlon, which requires participants to attempt ten separate events including sprints, jumps, vaults, and throwing events over the course of two grueling days, are scored by adding the tallies from each separate event to determine a final score. In the same way, track team scores are found by adding the scores from each individual event, relay, and field event to determine a total score.

Although the sport of bowling is scored using only addition, this popular game has one of the more unusual scoring systems in modern sports. Bowlers compete in games consisting of ten frames, each of which includes up to two attempts to knock down all ten bowling pins. Depending on a bowler's performance in one frame, he may be able to add some shots twice, significantly raising his total score. For example, a bowler who knocks down all ten pins in a single roll is awarded a strike, worth ten plus the total of the next two balls bowled in the following frames, while a bowler who knocks down all ten pins in two rolls is scored a spare and receives ten plus the next one ball rolled. Without this scoring system, the maximum bowling score would be earned by bowling ten, ten-point strikes in a row for a perfect game total of 100. But with bowling's bonus scoring system, each of the ten frames is potentially worth thirty points to a bowler who bowls a strike followed by two more strikes, creating a maximum possible game score of 300.

While many programs exist to help people lose weight, none is more basic, or less liked, than the straightforward process of counting calories. Calorie counting is based on a simple, immutable principle of physics: if a human body consumes more calories than it burns, it will store the excess calories as fat, and will become heavier. For this reason, most weight loss plans address, at least to some degree, the number of calories being consumed. A calorie is a measure of energy, and 3,500 calories are required to produce one pound of body weight. Using simple addition, it becomes clear that eating an extra 500 calories per day will add up to 3,500 calories, or one pound gained, per week.

While this use of addition allows one to calculate the waistline impact of an additional dessert or several soft drinks, a similar process defines the amount of exercise required to lose this same amount of weight. For example, over the course of a week, a man might engage in a variety of physical activities, including an hour of vigorous tennis, an hour of slow jogging, one hour of swimming, and one hour officiating a basketball game. Each of these activities burns calories at a different rate. Using a chart of calorie burn rates, we determine that tennis burns 563 calories per hour, jogging burns 493 calories per hour, swimming burns 704 calories per hour, and officiating a basketball game burns 512. Adding these values up we find that the man has exercised enough to burn a total of 2,272 calories over the course of the week. Depending on how many calories he consumes, this may be adequate to maintain his weight. However if he is

consuming an extra 3,500 calories per week, he will need to burn an additional 1,228 calories to avoid storing these extra calories as fat. Over the course of a year, this excess of 1,228 calories will eventually add up to a net gain of more than 63,000 calories, or a weight gain of more than 18 pounds.

While healthy activities help prolong life, the same result can be achieved by reducing unhealthy activities. Cigarette smoking is one of the more common behaviors believed to reduce life expectancy. While most smokers believe they would be healthier if they quit, and cigarette companies openly admit the dangers of their product, placing a health value or cost on a single cigarette can be difficult. A recent study published in the British Medical Journal tried to estimate the actual cost, in terms of reduced life expectancy, of each cigarette smoked. While this calculation is admittedly crude, the study concluded that each cigarette smoked reduces average life-span by eleven minutes, meaning that a smoker who puffs through all 20 cigarettes in a typical pack can simply add up the minutes to find that he has reduced his life expectancy by 220 minutes, or almost four hours. Simple addition also tells him that his pack-a-day habit is costing him 110 hours of life for each month he continues, or about four and one-half days of life lost for each month of smoking. When added up over a lifetime, the study concluded that smokers typically die more than six years earlier than non-smokers, a result of adding up the seemingly small effects of each individual cigarette.

FINANCIAL ADDITION

One of the more common uses of addition is in the popular pastime of shopping. Most adults understand that the price listed on an item's price-tag is not always the full amount they will pay. For example, most states charge sales tax, meaning that a shopper with $20.00 to spend will need to add some set percentage to his item total in order to be sure he stays under budget and doesn't come up short at the checkout counter. Many people estimate this add-on unconsciously, and in most cases, the amount added is relatively small.

In the case of buying a car, however, various add-ons can quickly raise the total bill, as well as the monthly payments. While paying 7% sales tax on a $3.00 purchase adds only twenty-one cents to the total, paying this same flat rate on a $30,000 automobile adds $2,100 to the bill. In addition, a car purchased at a dealership will invariably include a lengthy list of additional items such as documentation fees, title fees, and delivery charges, which must be added to the sticker price to determine the actual cost to the buyer.

As of 1999, Americans spent almost 40 cents of every food dollar at the 300,000 fast food restaurants in the country. Because they are often in a hurry to order, many customers choose one of the so-called value meals offered at most outlets. But in some cases, simple addition demonstrates that the actual savings gained by ordering a value meal is only a few cents. By adding the separate costs of the individual items in the meal, the customer can compare this total to learn just how much he is saving. He can also use this simple addition to make other choices, such as substituting a smaller order of French fries for the enormous order usually included or choosing a small soda or water in place of a large drink. Because most customers order habitually, few actually know the value of what they are receiving in their value meals, and many could save money by buying *à la carte* (piece by piece).

Deciding whether to fly or to drive is often based on cost, such as when a family of six elects to drive to their vacation destination rather than purchasing six airline tickets. In other cases, such as when a couple in Los Angeles visits relatives in Connecticut over spring break, the choice is motivated by sheer distance. But in some situations, the question is less clear, and some simple addition may reveal that the seemingly obvious choice is not actually superior. Consider a student living in rural Oklahoma who wishes to visit his family in St. Louis. This student knows from experience that driving home will take him eight hours, so he is enthusiastic about cutting that time significantly by flying. But as he begins adding up the individual parts of the travel equation, he realizes the difference is not as large as he initially thought. The actual flight time from Tulsa to St. Louis is just over one hour, but the only flight with seats available stops in Kansas City, where he will have to layover for two hours, making his total trip time from Tulsa to St. Louis more than three hours. Added to this travel time is the one hour trip from his home to the Tulsa airport, the one hour early he is required to check in, the half hour he will spend in St. Louis collecting his baggage and walking to the car, and the hour he will spend driving in St. Louis traffic to his family's home. Assuming no weather delays occur and all his flight arrive on time, the student can expect to spend close to seven hours on his trip, a net savings of one hour over his expected driving time. Simple addition can help this student decide whether the price of the plane ticket is worth the one hour of time saved.

In the still-developing world of online commerce, many web pages use an ancient method of gauging popularity: counting attendance. At the bottom of many web pages is a web counter, sometimes informing the

visitor, "You are guest number . . .". While computer gurus still hotly debate the accuracy of such counts, they are a common feature on websites, providing a simple assessment of how many guests visited a particular site.

In some cases, simple addition is used to make a political point. Because the United States government finances much of its operations using borrowed money, concerns are frequently raised about the rapidly rising level of the national debt. In 1989, New York businessman Seymour Durst decided to draw attention to the spiraling level of public debt by erecting a National Debt Clock one block from Times Square. This illuminated billboard provided a continuously updated total of the national debt, as well as a sub-heading detailing each family's individual share of the total. During most of the clock's lifetime, the national debt climbed so quickly that the last digits on the counter were simply a blur. The clock ran continuously from 1989 until the year 2000, when federal budget surpluses began to reduce the $5.5 trillion debt, and the clock was turned off. But two years later, with federal borrowing on the rise once again, Durst's son restarted the clock, which displayed a national debt of over $6 trillion. By early 2005, the national debt was approaching $8 trillion.

POKER, PROBABILITY, AND OTHER USES OF ADDITION

While predicting the future remains difficult even for professionals such as economists and meteorologists, addition provides a method to make educated guesses about which events are more or less likely to occur. Probability is the process of determining how likely an event is to transpire, given the total number of possible outcomes. A simple illustration involves the roll of a single die; the probability of rolling the value three is found by adding up all the possible outcomes, which in this case would be 1, 2, 3, 4, 5, or 6 for a total of six possible outcomes. By adding up all the possibilities, we are able to determine that the chance of rolling a three is one chance in six, meaning that over many rolls of the die, the value three would come up about 1/6 of the time. While this type of calculation is hardly useful for a process with only six possible outcomes, more complex systems lend themselves well to probabilistic analysis. Poker is a card game with an almost infinite number of variations in rules and procedures. But whichever set of rules is in play, the basic objective is simple: to take and discard cards such that a superior hand is created. Probability theory provides several insights into how poker strategy can be applied.

Consider a poker player who has three Jacks and is still to be dealt her final card. What chance does she have of receiving the last Jack? Probability theory will first add up the total number of cards still in the dealer's stack, which for this example is 40. Assuming the final Jack has not been dealt to another player and is actually in the stack, her chance of being dealt the card she wants is 1 in 40. Other situations require more complex calculations, but are based on the same process. For example, a player with two pair might wonder what his chance is of drawing a card to match either pair, producing a hand known as a full house. Since a card matching either pair would produce the full house, and since there are four cards in the stack which would produce this outcome, the odds of drawing one of the needed cards is now better than in the previous example. Once again assuming that 40 cards remain in the dealer's stack and that the four possible cards are all still available to be dealt, the odds now improve to 4 in 40, or 1 in 10. Experienced poker players have a solid grasp of the likelihood of completing any given hand, allowing them to wager accordingly.

Probability theory is frequently used to answer questions regarding death, specifically how likely one is to die due to a specific cause. Numerous studies have examined how and why humans die, with sometimes surprising findings. One study, published by the National Safety Council, compiled data collected by the National Center for Health Statistics and the U.S. Census Bureau to predict how likely an American is to die from one of several specific causes including accidents or injury. These statistics from 2001 offer some insight into how Americans sometimes die, as well as some reassurance regarding unlikely methods of meeting one's end.

Not surprisingly, many people die each year in transportation-related accidents, but some methods of transportation are much safer than others. For example, the lifetime odds of dying in an automobile accident are 1 in 247, while the odds of dying in a bus are far lower, around 1 in 99,000. In comparison, other types of accidents are actually far less likely; for instance, the odds of being killed in a fireworks-related accident are only 1 in 615,000, and the odds of dying due to dog bites is 1 in 147,000. Some types of accidents seem unlikely, but are actually far more probable than these. For example, more than 300 people die each year by drowning in the bathtub, making the lifetimes odds of this seemingly unlikely demise a surprising 1 in 11,000. Yet the odds of choking to death on something other than food are higher by a factor of ten, at 1 in 1,200, and about the same as the odds of dying in a structural fire (1 in 1,400) or being poisoned (1 in 1,300). Unfortunately, these odds are roughly equivalent to the lifetime chance of dying due to medical or surgical errors or complications, which is calculated at 1 in 1,200.

Geometric Progression

An ancient story illustrates the power of a geometric progression. This story has been retold in numerous versions and as taking place in many different locales, but the general plot is always the same. A king wishes to reward a man, and the man asks for a seemingly insignificant sum: taking a standard chessboard, he asks the king to give him one grain of rice on day one, two grains of rice on day two, and so on for 64 days. The king hastily agrees, not realizing that in order to provide the amount of rice required he will eventually bankrupt himself.

How much rice did the king's reward require? Assuming he could actually reach the final square of the board, he would be required to provide 9,223,372,036,854,775,808 grains of rice, which by one calculation could be grown only by planting the entire surface of the planet Earth with rice four times over. However it is doubtful the king would have moved far past the middle section of the chessboard before realizing the folly of his generosity. The legend does not record whether the king was impressed or angered by this demonstration of mathematical wisdom.

USING ADDITION TO PREDICT AND ENTERTAIN

Addition can be used to predict future events and outcomes, though in many cases the results are less accurate than one might hope. For example, many children wonder how tall they will eventually become. Although numerous factors such as nutrition and environment impact a person's adult height, a reasonable prediction is that a boy will grow to a height similar to that of his father, while a girl will approach the height of her mother. One formula which is sometimes used to predict adult height consists of the following: for men, add the father's height, the mother's height, and 5, then divide the sum by 2. For women, the formula is (father's height + mother's height − 5) / 2. In most cases, this formula will give the expected adult height within a few inches.

One peculiar application of addition involves taking a value and adding that value to itself, then repeating this operation with the result, and so forth. This process, which doubles the total at each step, is called a geometric progression, and beginning with a value of 1 would appear as 1, 2, 4, 8, 16, 32 and so forth. Geometric progressions are unusual in that they increase very slowly at first, then more rapidly until in many cases, the system involved simply collapses under the weight of the total.

One peculiarity of a geometric progression is that at any point in the sequence, the most recent value is greater than the sum of all previous values; in the case of the simple progression 1, 2, 4, 8, 16, 32, 64, addition demonstrates that all the values through 32, when added, total only 63, a pattern which continues throughout the series. One seemingly useful application of this principle involves gambling games such as roulette. According to legend, an eighteenth century gambler devised a system for casino play which used a geometric progression. Recognizing that he could theoretically cover all his previous losses in a single play by doubling his next bet, he bragged widely to his friends about his method before setting out to fleece a casino. The gambler's system, known today as the Martingale, was theoretically perfect, assuming that he had adequate funds to continue doubling his bets indefinitely. But because the amount required to stay in the game climbs so rapidly, the gambler quickly found himself out of funds and deep in debt. While the story ends badly, the system is mathematically workable, assuming a gambler has enough resources to continue doubling his wagers. To prevent this, casinos today enforce table limits, which restrict the maximum amount of a bet at any given table.

Addition also allows one to interpret the cryptic-looking string of characters often seen at the end of series of motion picture credits, typically something like "Copyright MCMXXLI." While the modern Western numbering system is based on Arabic numerals (0–9), the Roman system used a completely different set of characters, as well as a different form of notation which requires addition in order to decipher a value. Roman numerals are written using only seven characters, listed here with their corresponding Arabic values: M (1,000), D (500), C (100), L (50), X (10), V (5), and I (1). Each of these values can be written alone or in combination, according to a set of specific rules. First, as long as characters are placed in descending order, they are simply added to find the total; examples include VI (5 + 1 = 6), MCL (1,000 + 100 + 50 = 1,150), and LIII (50 + 1 + 1 + 1 = 53). Second, no more than two of any symbol may appear consecutively, so values such as XXXX and MCCCCV would be incorrectly written.

Because these two rules are unable to produce certain values (such as 4 and 900), a third rule exists to handle these values: any symbol placed out of order in the descending sequence is not added, but is instead subtracted from the following value. In this way, the proper sequence for 4 may be written as IV (1 subtracted from 5), and the

An Iraqi election officer checks ballot boxes at a counting center in Amman, Jordan, 2005. Counting ballots was accomplished by adding ballots one at a time, by hand. AP/WIDE WORLD PHOTOS. REPRODUCED BY PERMISSION.

Roman numeral for 900 is written CM (100 subtracted from 1,000). While this process works well for shorter numbers, it becomes tedious for longer values such as 1997, which is written MCMXCVII (1,000 − 100 + 1,000 − 10 + 100 + 5 + 1 + 1). Adding and multiplying Roman numerals can also become difficult, and most ancient Romans were skilled at using an abacus for this purpose. Other limitations of the system include its lack of notation for fractions and its inability to represent values larger than 1,000,000, which was signified by an M with a horizontal bar over the top. For these and other reasons, Roman numerals are used today largely for ornamental purposes, such as on decorative clocks and diplomas.

Potential Applications

While addition as a process remains unchanged from the method used by the ancient Chinese, the mathematical tools and applications related to it continue to evolve. In particular, the exponential growth of computing power will continue to radically alter a variety of processes. Gordon Moore, a pioneer in microprocessor design, is credited with the observation that the number of transistors on a processor generally doubles every two years; in practice, this advance means that computer processing power also doubles. Because this trend follows the principle of the geometric progression, with its doubling of size at each step, expanding computer power will create unexpected changes in many fields. As an example, encryption schemes, which may use a key consisting of 100 or more digits to encode and protect data, could potentially become easily decipherable as computer power increases. The rapid growth of computing power also holds the potential to produce currently unimaginable applications in the relatively near future. If the consistent geometric progression of Moore's law holds true computers one decade in the future will be fully 32 times as powerful as today's fastest machines.

Where to Learn More

Books

Orkin, Mike. *What are the Odds? Chance in Everyday Life*. New York: W.H. Freedman and Company, 2000.

Seiter, Charles. *Everyday Math for Dummies*. Indianapolis: Wiley Publishing, 1995.

Walker, Roger S. *Understanding Computer Science*. Indianapolis: Howare W. Sams & Co., 1984.

Periodicals

Lin, B-H., E. Frazao, and J. Guthrie. "Away-From-Home Foods Increasingly Important to Quality of American Diet," *Agricultural Information Bulletin, U.S. Department of Agriculture and U.S. Department of Health and Human Services*. (1999).

Shaw, Mary, Richard Mitchell, and Danny Dorling. "Time for a smoke? One cigarette reduces your life by 11 minutes." *British Medical Journal*. (2000): 320 (53).

Web sites

Aetna Intellihealth. "Can We Predict Height?" February 11, 2003. <http://www.intelihealth.com/IH/ihtIH/WSIHW000/35320/35323/360788.html?d=dmtHMSContent#bottom> (March 15, 2005).

Arabic 2000. "The Arabic Alphabet." <http://www.arabic2000.com/arabic/alphabet.html> (March 15, 2005).

Brillig.com. "U.S. National Debt Clock." <http://www.brillig.com/debtclock> (March 15, 2005).

ComputerWorld. *"Inside a Microprocessor."* <http://www.computerworld.com/hardwaretopics/hardware/story/0,10801,64489,00.html> (March 14, 2005).

DECA: The Decathlon Association. "The Decathlon Rules." <http://www.decathlonusa.org/rules.html> (March 15, 2005).

Gambling Gates. "The truth behind the limits: What maximum and minimum bets are about." <http://www.gamblinggates.com/Tips/maximum_bets8448.html> (March 15, 2005).

Intel Research. "Moore's Law." <http://www.intel.com/research/silicon/mooreslaw.htm> (March 15, 2005).

Mathematics Magazine. "Chisenbop Tutorial." <http://www. mathematicsmagazine.com/5-2003/Chisenbop_5_2003.htm> (March 13, 2005).

National Safety Council. "What are the odds of dying?" <http://www.nsc.org/lrs/statinfo/odds.htm> (March 15, 2005).

Nutristrategy. "Calories Burned During Exercise." <http://www.nutristrategy.com/activitylist.htm> (March 15, 2005).

Sigma Educational Supply. "History of the Abacus." <http://www.citivu.com/usa/sigmaed/> (March 14, 2005).

Tallahassee Democrat (AP). "National 'Debt Clock' Restarted." July 11, 2002. <http://www.tallahassee.com/mld/tallahassee/business/3643411.htm> (March 15, 2005).

The Great Idea Finder. "Fascinating Facts About the Invention of the Abacus by Chinese in 3000 BC." Inventions. <http://www.ideafinder.com/history/inventions/abacus.htm> (March 14, 2005).

The Math Forum at Drexel. "How Are Roman Numerals Used today?" (March 15, 2005).

University of Maryland Physics Department. "A2-61: Exponential Increase - Chessboard and Rice." <http://www.physics.umd.edu/lecdem/services/demos/demosa2/a2-61.htm> (March 15, 2005).

Algebra

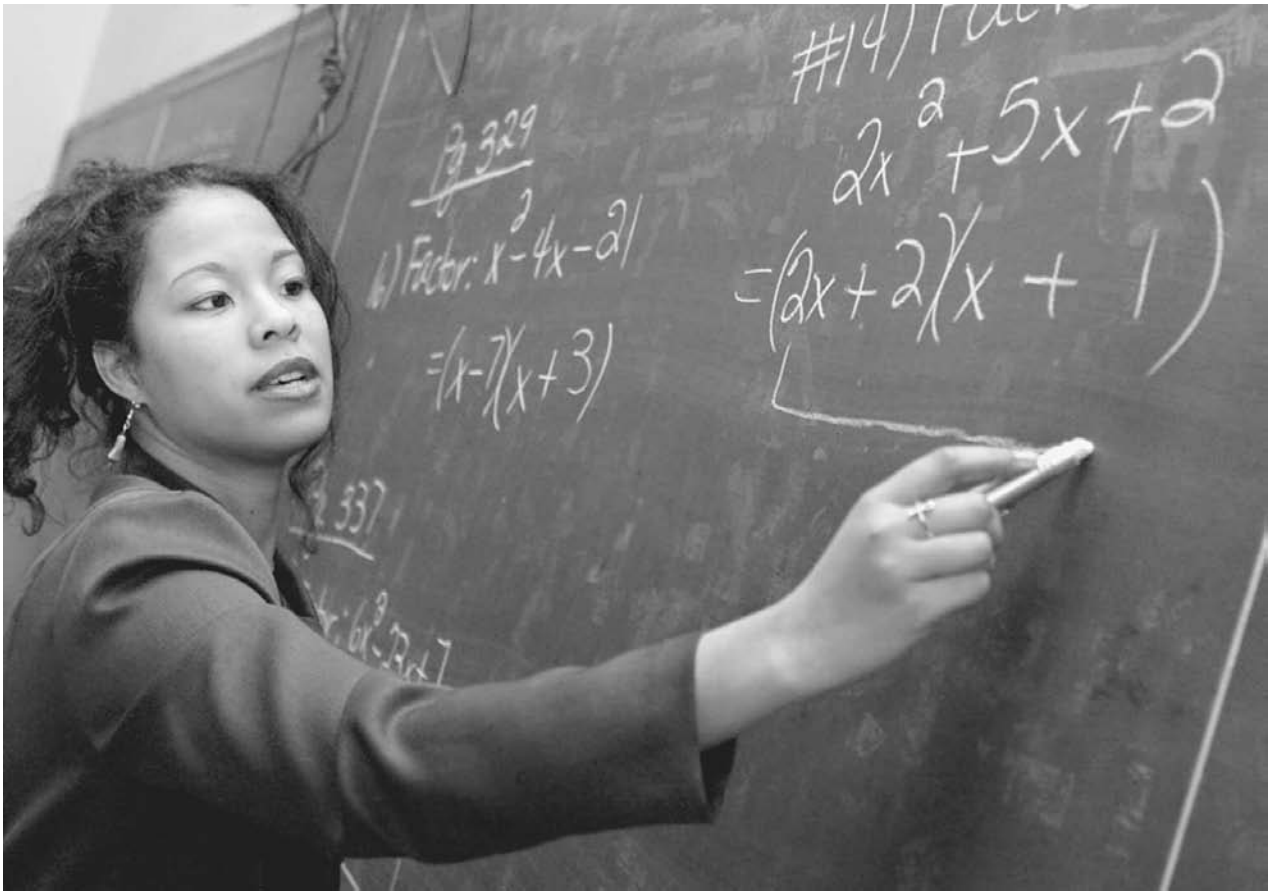
Overview

Algebra is the study of mathematical procedures that combine basic arithmetic with a wide range of symbols in order to express quantitative concepts. Arithmetic refers to the study of the basic mathematical operations performed on numbers, including addition, subtraction, multiplication, and division, and is widely viewed as a separate field of mathematics because it must be taught to students before they can progress to higher studies; but arithmetic is basically algebra without the symbols and advanced operations. In this sense, algebra is often referred to as a generalization of arithmetic, which can be applied to more sophisticated ideas than numbers alone. From adding up the price of groceries and balancing a checkbook, to preparing medicines or launching humans into space, algebra enables almost any idea to be written in standard mathematical notation that can be utilized by people around the world. No matter how advanced the mathematics involved, algebraic rules and notations provide the instructions that dictate how to handle the various combinations of numbers and symbols.

Fundamental Mathematical Concepts and Terms

Algebraic symbols can be classified into symbols for representing quantities (usually numbers and letters); symbols for representing operations (such as addition, subtraction, multiplication, division, exponents, and roots); symbols representing equality and inequality (equal to, approximately equal to, less than, greater than, less than or equal to, greater than or equal to, and not equal to); and symbols for separating and organizing terms, and determining the order of operations (typically parentheses and brackets).

Multiplication in an algebraic expression is often represented by a dot when written out by hand (e.g., $4 \cdot 5 = 20$), or an asterisk when using a computer or graphing calculator (e.g., $4*5 = 20$). Adjacent sets of parentheses also signify multiplication, as in $(4)(5) = 20$. A number or variable attached to the outside of a set of parentheses also signifies multiplication. That is, $60 \times t = (60)(t) = (60)t = 60(t)$. These notations are used instead of 4×5 and $60 \times t$ in order to avoid confusion between the multiplication sign and the commonly used variable x. The symbol for multiplication is often omitted from an equation altogether: aside from the notation, $60t$ is identical to $60 \times t$. When two numbers are multiplied together, there must always be some sort of symbol to



Mathematician Dr. Tasha Inniss corrects a factorization shown on blackboard. Can you spot the error? AP/WIDE WORLD PHOTOS. REPRODUCED BY PERMISSION.

indicate the multiplication in order to avoid confusion. For example, $(2)(3) = 2 \times 3 \neq 23$.

Repeated multiplication can be simplified using exponential notation. If the letter n is used to represent a generic number, then $n \times n = n^2$ (n squared), $n \times n \times n = n^3$ (n cubed or n to the 3rd power), and so on. For example, if $n = 3$, then $n^3 = 3^3 = 3 \times 3 \times 3 = 27$. In general, the value of n multiplied by itself y times can be expressed as n^y, read n to the power of y.

Performing operations in the proper order is essential to finding the correct solution to an equation. In general, the proper order of operations is as follows:

1. Using the following guidelines, always perform operations moving from left to right;
2. Perform operations within parentheses or brackets first;
3. Next, evaluate exponents;
4. Then perform multiplication and division operations;
5. Finally, perform addition and subtraction operations.

In algebraic equations, numbers are typically referred to as constants because their values do not change. Letters are most often used as variables, which represent either unknown values or placeholders that can be replaced with any value from a range of numbers. For example, if a car is traveling at a speed of 60 miles per hour, then the distance that the car has traveled can be represented as $d = 60t$, where d represents the distance in miles that the car has traveled and t represents the number of hours that the car has been moving. The variable t can be replaced with any nonnegative value (zero and the positive numbers); as time progresses, t increases, and as would be expected, the distance d increases.

An expression that involves variables, numbers, and operations is called a variable expression, or algebraic expression. For example, $x^2 + 3x$ is a variable expression. An equation, like $x^2 + 3x = 18$, is created when a variable expression is set equal to a number, variable, or another variable expression. An algebraic inequality is expressed

when a variable expression is separated from a number, variable, or another variable expression by a greater than sign, less than sign, greater than or equal to sign, or less than or equal to sign. Inequalities can be used to determine upper or lower bounds for a possible range of values. For example, the idea that it takes less than 15 minutes to boil an egg can be expressed as $t < 15$, where t represents time measured in minutes.

The parts of an equation that are separated by the symbols of addition, subtraction, and equality (or inequality) are called the terms of the equation. In the equation $4x^2 + 3x = 76$, the three terms are $4x^2$, $3x$, and 76. The symbols of positive and negative can also be taken into account in the terms of the equation, so that the terms are only separated by the symbols of addition and equality. In the equation $8x^2 - 3x = 26$, the terms are $8x^2$, $-3x$, and 26, because $8x^2 - 3x$ can be written as $8x^2 + (-3x)$. In general, subtraction can be thought of as addition of a negative term.

When a constant and a variable are multiplied, the constant is called the coefficient of the term. In the variable expression $8x^2 - 3x$, the coefficient of the first term is 8 and the coefficient of the second term is -3.

A special type of equation or inequality in which there are an infinite number of solutions is known as an algebraic formula. Formulas are useful for performing repeated mathematical tasks. The previous equation for determining the distance that a car has traveled if traveling at 60 miles per hour for a given amount of time, $d = 60t$, is a formula because for every value of t, there is a new value for d. If the value for d is known, the value of t can be determined, and vice versa. This formula can be generalized to allow for different speeds as well. In the formula $d = st$, any speed can be substituted for the variable s. An equation like $2x^2 + x = 10$ is not a formula because only a finite number of values of x satisfy the equation.

Equations in which the highest power of any term is one are called linear equations (recall that $x^1 = x$). The equation $d = 60t$, for instance, is linear. Nonlinear equations are those that involve at least one term raised to a power greater than one. Equations in which the highest power of any term is two are referred to as quadratic equations. The equation $5x^2 + 3x = 2$ is an example of a quadratic equation. In general, a quadratic equation can be simplified into the form $ax^2 + bx + c = 0$, where a, b, and c are the coefficients of the terms. There are various methods for solving quadratic equations. One of the most common methods is the known as the quadratic formula, which states that

$$x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$$

For example, the equation $5x^2 + 3x = -2$ can be rewritten as $5x^2 + 3x + 2 = 0$ (by adding 2 to both sides of the equation); so the values of the coefficients are $a = 5$, $b = 3$, and $c = 2$. Substituting these values into the quadratic formula reveals the values of x that satisfy the equation:

$$x = \frac{-3 \pm \sqrt{3^2 - 4(5)(-2)}}{2(5)} = \frac{-3 \pm \sqrt{9 + 40}}{10} = \frac{-3 \pm 7}{10}$$

Therefore, the values of x that satisfy this equation are $-\frac{2}{5}$ and -1. These values can be substituted for x to verify that they satisfy the equation.

Equations in which the highest power of any term is three are called cubic equations. Equations involving higher powers are usually referred to as 4th-order equations, 5th-order equations, and so on.

The various methods and rules for simplifying the terms of an algebraic equation constitute the most important tools for working with any mathematical construction. For example, rules like the associative, commutative, and distributive properties dictate how terms can be added and multiplied to simplify and solve algebraic expressions.

Combining like terms is a useful method of simplification. To illustrate this method, consider the task of counting the number of boys and girls in a gymnasium. One way to simplify this problem is to ask all of the boys to move to one side of the room, and all the girls to move to the other side. Similar reasoning is used to simplify a messy algebraic equation like $3x^2 - 7x - x^2 + 9x + x^2 - 4x - 4x^2 + 2x^2 - 2 = 6$. Terms involving the same variable raised to the same power are called like terms and can be added and subtracted just like numbers. This equation involves three powers of the variable x, so by collecting like terms it can be simplified to an equation with only three terms. First, by grouping the like terms together, the equation becomes $3x^2 - x^2 + x^2 - 4x^2 + 2x^2 - 7x + 9x - 4x = 6 + 2$ (note that 2 was added to each side of the equation in order to group the constants on the right side). Next, by adding and subtracting like terms, the equation becomes less of any eyesore: $x^2 - 2x = 8$.

Factoring allows seemingly difficult equations to be expressed in different ways that can immediately reveal a solution. For example, finding the values of x that satisfy the equation $x^2 - 2x = 8$ may at first seem intimidating; but this equation can be rewritten as $(x - 4)(x + 2) = 0$, which reveals that $x = 4$ and $x = -2$ both satisfy the equation (if either of these values is substituted for x, then one of the two parenthetical expressions is equal to

zero, so when it is multiplied with the other expression, the entire left side is equal to zero).

The endless rules and tricks for evaluating algebraic expressions permeate mathematics and science at all levels. Whether noticed not, the fundamental concepts of algebra can be found in daily activities, and can be used to explain many concepts in the universe.

A Brief History of Discovery and Development

The symbols and syntax of algebra have developed slowly over thousands of years to become what is now recognized as the fundamental language of mathematics. In ancient times, mathematical problems were often written out in the verbal language of the time. As similar problems repeatedly arose, people began to invent abbreviations, and eventually symbols, for common terms in the problems. As mathematical concepts progressed and great mathematicians continued to make breakthroughs based on the findings of earlier mathematicians, less words were used to describe problems and the language of mathematics was continuously refined and adapted to the pressing problems of the various times and civilizations. Eventually, almost any problem or arithmetical fact that humans found could be expressed using the mathematical symbols of algebra.

In general, algebra refers to mathematical operations involving unknown values represented by some sort of symbols, where other symbols are used as shorthand for commands. In this sense, algebra was studied extensively in ancient Egypt, possibly as early as 2000 B.C. However, algebra in the form that is recognized today (even the word algebra) would not be discovered for thousands of years after the Egyptians began using these concepts. In ancient Egypt, and in many civilizations prior to the development of the current conception of algebra, algebraic equations were not generalized as ideas that could be applied to other types of problems. Individual practical problems of the time were studied and documented. Once a problem was solved, the writings could be used to solve a similar problem using different values for the unknowns; but mathematical language was seldom shared between different types of problems. For example, the method for finding the optimal amount of fertilizer to place on a crop was not seen as related mathematically to the method for figuring how much grain would fit in a storage structure, even though both procedures involve the operations now known as addition and multiplication.

The Greek mathematician Diophantus made great progress in generalizing algebraic symbolism in his writings. Little is known of Diophantus' life, but it is commonly held that his most important works took place about A.D. 250. He discovered a general method for solving equations and finding values of unknowns. Diophantus is attributed as the first mathematician to use abbreviations for unknowns (variables) and powers of unknowns, and abbreviating words such as the Greek word meaning "is equal to". The use of these abbreviations was a major step toward the sophisticated algebraic symbolism (e.g., using letters to represent variables) found in modern mathematics. However, Diophantus did not use notation that could represent two or more unknowns and resorted to using words to describe multiple unknowns.

Like the earlier Egyptians, Diophantus viewed his mathematical ideas not as theories in the workings of numbers, but as a means for solving common problems of his day. His main work, *Arithmetica*, is a collection of pertinent problems described using numerical solutions of mathematical structures, the predecessors of algebraic equations. Diophantus' original works did not compensate for abstract ideas such as negative numbers. The idea of a negative number, or an equation like $x + 20 = 2$, was not explored because the idea of a negative quantity, a negative stone or book for example, was not needed in his society. Nonetheless, an essential branch of algebraic analysis that deals with solving certain types of rational equations (equations that allow for fractions and roots in addition to whole numbers) has been named Diophantine analysis (or analysis of Diophantine equations) in celebration of his work.

An Arab mathematician named Abu Abdullah Muhammad ibn Musa al-Khwarizmi contributed greatly to the language and concepts of algebra. Like Diophantus, little is known about the life of Khawarizmi, but it is commonly accepted that most of his important works took place around A.D. 820. In addition to his resounding developments in various fields of mathematics, he also contributed greatly to astronomy, geography, the inner workings of clocks, and the degree measurements of angles. Khawarizmi's writings on arithmetic and algebra have had resounding effects on the fundamental ideas of modern mathematics.

Based on knowledge documented by Greek mathematicians and the innovative notation for numbers and mathematical operations proposed by Hindu contemporaries in India, Khawarizmi developed the basis for modern arithmetical notation. His writings introduced and developed several fundamental arithmetic procedures, including operations performed on fractions. He was the first to spread the decimal number system (now commonly referred to as Arabic numerals) and the idea of the

number zero outside of India, introducing it directly to Arabs, and later to Europe when his writings were translated to Latin and other European languages. Khawarizmi's original book on arithmetic was lost, leaving only translations.

Another of Khawarizmi's books, *Kitab al-Jabr w'al-Muqabala*, sparked the analysis of algebra as a well-organized form of mathematics. The title of the book has been interpreted in various ways, including "Rules of Reintegration and Reduction" and "The book of summary concerning calculating by transposition and reduction." The word algebra is derived from the term al-jabr in the title of the book, which can be taken to mean "reunion of broken parts," "reduction," "connection," or "completion." The rest of the title loosely translates to "to set equal to" or "to balance." The title of the book relates to the fundamental procedures involved in solving algebraic problems, such as shifting terms from one side of an equation to the other and combining like terms. The methods described in Khawarizmi's books have been built upon ever since; and the word algebra evolved for centuries before it was spelled and used as it is today.

At the beginning of the thirteenth century, Leonardo da Pisa (also known as Leonardo Fibonacci), an Italian mathematician, traveler, and tradesman, discovered that the potential of algebraic computations using the Hindu (Arabic) notation for numbers far exceeded the capacities of the Roman numeral system that was standard in Europe at that time. In his writings on algebra, he discussed the superiority of the symbols and concepts borne in distant lands. His writings included little original discoveries and were intended to illuminate pertinent ideas and problems found in his culture at that time. Unfortunately, his proposals were generally viewed as nothing more than interesting, and the ideas that he attempted to spread would not catch on in Europe for almost 300 years.

In the late fifteenth century, an Italian named Lucas Paciolus (Lucas de Burgo) authored multiple works on arithmetic, geometry, and algebra. Though most of the mathematical elements are taken from earlier writings, his algebraic writings were integral in the development of algebraic methods because of his efficient use of symbols. Due to the invention of the printing press earlier in the century, Paciolus' writings were among the first widely distributed algebraic texts, at long last effectively introducing the benefits of algebraic reasoning and Arabic numerals.

In the sixteenth century, algebra began to be used in a purely mathematical sense, with symbols and numbers completely representing general quantitative ideas. Robert Recorde—an English mathematician and originator of the symbol = for representing equality—is attributed with the first use of the term algebra in a strictly mathematical sense. François Viete made much progress in the use of symbols for representing generic numbers, which enabled mathematic ideas to be represented in a more general manner and ultimately led to the view of algebra as generalized arithmetic.

The recognition and understanding of negative values, irrational numbers, and negative roots of quadratic equations were crucial developments in the progression of algebraic theories because they opened doors to more advanced mathematical concepts. The discovery of negative numbers is often attributed to Albert Girard in the early seventeenth century. Unfortunately for Girard, the work of René Descartes—another great mathematician of the time—overshadowed his findings.

In the field of algebra, the most notable accomplishment of Descartes was the discovery of relationships between geometric measurements and algebraic methods, now referred to as analytic geometry (geometry analyzed using algebra). Using this analytic method of describing measurements such as lengths and angles, Descartes showed that algebraic manipulations (e.g., addition, multiplication, extraction of roots, and representation of negative values) could be represented by investigating related geometric shapes. Descartes' fusion of algebra and geometry elucidated both mathematical fields.

In the more than two centuries following Descartes discoveries, mathematicians have continued to refine algebraic notation and analyze the properties of more sophisticated aspects of mathematics. Many algebraic advances enable mathematicians and scientists to investigate and understand real-world phenomena that were previously thought impossible or unnecessary to analyze. In the twenty-first century, it seems that there are as many types of algebra as there are problems to be solved, but all of them depend wholly on the concepts of basic algebra.

Real-life Applications

PERSONAL FINANCES

Many people use a checkbook registry or financial software to track their income and expenses in order to make sure that they are making enough money to pay their bills and accomplish their financial goals. The registry in a checkbook is basically a form that helps to perform the algebraic operations necessary to track expenses. A checkbook registry includes columns for describing transactions (including the dates on which

they take place), and columns for the recording amount of each transaction. There is usually a column labeled "deposits" and another column labeled "debits" so that transactions that add money and transactions that subtract money can be kept separate for quick and easy analysis. For example, when birthday money or a paycheck is received, it is logged in the deposits column. Things like groceries, rent, utilities, and car payments are recorded in the debits column. A "balance" column is provided for calculating the amount of money present in the bank account after each transaction. The process of recording transactions and balances in a checkbook registry basically involves performing a large, highly descriptive, ongoing algebraic equation. When a transaction is recorded in the column for deposits, a positive term is appended to the equation. When a transaction is recorded in the debits column, a negative term is appended to the equation. The balance column represents the other side of the equation. As the terms are appended to the equation, the balance column may be updated immediately, or the various transactions can be recorded and the total can be found later; but either way, the balance is always the result of the addition and subtraction of the terms represented by the values in the debits and deposits columns.

Every April, millions of United States citizens must analyze their financial income from the previous year in order to determine how much income tax they are required pay to federal and state government offices. Government taxes help pay for many social benefits, such as healthcare and social security. The Internal Revenue Service (IRS) provides various forms with step by step instructions for performing algebraic operations to calculate subtotals, and ultimately the amount of money that must be sent to the government. The various items on a tax form include the amount of taxes that are withheld from each paycheck; the amount of money taken home from each check after estimated taxes are deducted; items of personal worth such as savings, investments, and major possessions; and work-related expenses such as company lunches, office supplies, and utility bills. Many people receive money back from the government because the items representing expenses and taxes throughout the year add up to more than the total taxes due for the year. Some people end up owing taxes at the end of the year. Other people, such as self-employed workers, may or may not have taxes withheld from each paycheck. These people generally use a different type of IRS form and need to save money throughout the year to pay their taxes come tax time. Whatever form is used, the various items are added, subtracted, multiplied and divided just like the terms of an algebraic equation. In essence, an IRS tax form is an expanded algebraic equation, with the terms and operations written out as explicit, intuitive instructions. The variables are described with words and a blank line or box is provided for filling in the value of each variable.

COLLEGE FOOTBALL

Unlike other college sports, National Collegiate Athletic Association (NCAA) football does not hold a national tournament at the end of the season to determine which team is the year's national champion. Instead, a total of 25 bowl games are held throughout the country, pitting teams with winning records against each other. The Bowl Championship Series (BCS) consists of four of these bowl games: the Orange Bowl, the Fiesta Bowl, the Sugar Bowl, and the Rose Bowl. These four bowl games feature eight of the highest rated teams of the year, and each year a different bowl game is designated as the national championship game. An invitation to any BCS game guarantees that a school will receive a hefty sum of money at the end of the year. Winning a BCS game could bring in millions of dollars.

The mathematical formula used to figure out which teams make it to the BCS games (and which two teams will fight to be crowned the national champions) turns out to be a rather complicated application of statistical analysis; but algebra provides the backbone of the entire operation. Across the country every week, each team's BCS ratings are updated according to four major factors: Computer rankings, the difficulty of the team's schedule, opinion polls, and the team's total number of losses. Each of these four components yields a numerical value.

The computer rankings, for one, are determined by complex computer programs created by statisticians. Computer ranking programs crunch an enormous amount of statistical data, including numerical values representing a multitude of factors ranging from the score of the game, the number of turnovers, and each team's total yardage, to the location of the game and the effects of weather.

The difficulty of a team's schedule is also determined by algebraic equations with terms accounting for the difficulty of the team's own schedule and the difficulty of the schedules of the teams that they will play throughout the season.

There are two separate opinions polls: one involving national sports writers and broadcasters, and one involving a select group of football coaches. Each poll results in a numerical ranking for all of the teams. For each team, an average of a these two rankings determines their national opinion poll ranking.

A team's number of losses is the most straightforward factor. The number of losses is figured directly into the general mathematical model, and each loss throughout the season has a large effect on a team's overall ranking.

The four numerical values are added together to calculate the team's national ranking. The top two teams at the end of the regular season are invited to the national championship BCS game. However, the selection of the six teams that are invited to the other three BCS games is not as straight forward. These other six teams are selected from the top 12 teams across the nation (excluding the top two that are automatically invited to the championship game). How these 12 teams are narrowed to six depends mainly on which teams are expected to attract the most attention and, therefore, create the most profits for the hosting institution, the television and radio stations that broadcast the game, and the various sponsors. These financial considerations are also modeled using algebraic formulas.

UPC BARCODES

Universal Product Code (UPC) barcodes are attached to almost all items purchased from mass merchandisers, such as department stores and grocery stores. These barcodes were originally used in grocery stores to help track inventory and speed up transactions, but shortly thereafter, UPC barcodes were appearing on all types of retail products.

UPC barcodes have two components: the barcode consisting of vertical lines that can be read by special scanning devices, and a set of numbers that can be read by humans (see Figure 1). Each component represents the same 12-digit number in a different language. That is, the barcode is simply the number below it represented in the language that can be read by the barcode scanner. The language of barcode scanners is based on vertical lines of two different colors (usually black and white) and four different sizes (the skinniest lines, and lines that are two, three, and four times as thick).

The UPC numbers for all items throughout the world are created and maintained by a central group called the Uniform Code Council (UCC). The first six numbers of a product's UPC number identify the manufacturer. Any manufacturer that wants to use UPC barcodes must submit an application to the UCC and pay an annual fee. Every barcode found on products sold by the same manufacturer will start with the same six digits. The first digit of the manufacturer number (the first digit in the entire UPC number) organizes all manufacturers into different categories. For example, the UPC numbers for

Figure 1: UPC bar code. KELLY QUIN. REPRODUCED BY PERMISSION.

pharmaceutical items, such as medicines and soaps, begin with 3. Some numbers at the beginning of UPC numbers are reserved for special items like coupons and gift certificates.

The second set of five digits represents the product itself. This five-digit product code is unique on every different product sold by a manufacturer, even different sizes of the same product. Some larger manufacturers have secured choice manufacturer codes and product codes that contain consecutive zeros. In certain configurations, consecutive zeros can be left out so that the UPC barcode can be squeezed onto small products, such as packs of chewing gum. There are ways to determine the positions of missing zeros when less than 12 numbers appear; but regardless, the barcode represents all 12 digits so that a quick swipe in front of a scanner determines all of the necessary information.

In any store, the price of each item is stored in a separate computer, which is attached to all of the checkout registers and provides the price for each item scanned. The prices of items are not indicated on barcodes because different stores charge different prices and all stores need to be able to change prices quickly.

The final digit of a UPC number is called the check digit and is used to minimize mistakes made by barcode scanners. The final digit can be calculated from the preceding 11 digits using a standard set algebraic operations. Following is an explanation of the algebra involved in calculating the check digit of the UPC number 43938200039, which has a check digit of 9:

1. Starting with the first digit, add together all of the digits in every other position (skipping every other number): $4 + 9 + 8 + 0 + 0 + 9 = 30$. In a sense, this sum is a variable in the equation for calculating the check digit because it represents values that can be changed.

2. Then multiply that value by 3: 3*30 = 90. The number 3 is a constant value in the check digit equation, and can be thought of as the coefficient of the variable in the previous step. Together, this coefficient and the variable sum in the previous step form a term in the equation for calculating the check digit.
3. Next, add up all of the other digits in the UPC number (starting with the second digit and skipping every other digit): 3 + 3 + 2 + 0 + 3 = 11. This sum is another term of the equation used to calculate the check digit. This variable value is not multiplied by a constant, so there is no coefficient of this term.
4. Now add the values of these two terms together: 90 + 11 = 101.
5. Finally, determine the smallest number that, when added to the value found in the previous step, results in a multiple of ten. That is, find the smallest number that can be added to the number in the previous step such that the sum divided by ten does not yield a remainder. In this case, that number is 9: 101 + 9 = 110. Using the mathematical concept of remainders, this value can be represented in an algebraic equation as well.
6. Compare the number found in the previous step with the final digit in the UPC number. The fact that this number matches the check digit in the UPC number confirms that the previous 11 digits were read correctly.

The entire calculation of the check digit can be represented by a single equation. A barcode scanner performs these calculations almost instantaneously every time a barcode is scanned. The actual check digit is represented at the end of the barcode (just as it appears at the end of the UPC number that humans can read). If the check digit calculated using the first 11 digits does not match the actual check digit, the scanner communicates to the store clerk—usually by making a loud beep and displaying a message on the screen of the cash register—that an error has occurred and the item needs to be rescanned.

FLYING AN AIRPLANE

In order for an 870,000-pound (394,625-kg) 747 jumbo jet to fly thousands of miles, it must be built according to strict specifications to create and balance the forces needed to carry this huge collection of metal through the air.

Thrust is the force that an airplane creates by moving forward very fast and causing air to move quickly past its wings. Airplanes use powerful propellers, jet engines, or rockets to create enough thrust to drive the airplane forward. When an airplane moves through the air, it also creates drag, a force that acts in the opposite direction of thrust and slows the plane down. When a hand is sticking out of a moving car, it creates similar drag. The faster the car is moving, the more the passing air acts on the hand, causing it to move backward with respect to the movement of the car. An airplane must create enough thrust to counteract the drag forces. This is why large planes must be traveling at hundreds of miles per hour in order to get off the ground. After a plane takes off, the landing gear is retracted because, much like a hand sticking out of a car, the landing gear creates drag. In fact, the drag created by the landing gear of a large airplane would most likely rip the landing gear off, leaving the pilots and passengers in a terrible predicament.

There are two other important forces that act on an airplane in motion. Gravity is pulling the airplane toward Earth so the weight (mass) of the airplane is an important factor. In order to raise the weight of the airplane upward, the airplane must create another force, called lift. As an over-simplified explanation of these four forces: the airplane must create enough thrust to move the plane quickly forward and counteract drag; and the airplane must be designed in such a way that when the plane moves forward fast enough, sufficient lift is created in order to counteract the forces of gravity acting on the mass of the plane. The wings play the biggest role in creating lift. The details of how an airplane creates lift involve advanced concepts of physics (including the idea that air is a fluid and acts much like water); but the calculations involved in planning and executing the safe operation of any airplane involve an immense amount of algebra. In algebraic equations that model the lift that an airplane produces, for instance, variables represent the factors that affect lift, including the density of the air, the speed of the airplane, the shape and surface area of the wings, and the angle at which the wings meet the oncoming air.

In addition to the calculations that must be checked and rechecked to ensure that an airplane can create sufficient lift, each trip involves a variety of important algebraic formulas. For example, deciding how much fuel to load into an airplane for each trip involves factors including the desired distanced to be traveled, the weather, and the effect that the weight of the fuel has on take off and landing procedures. Obviously, enough fuel must be present in the airplane to keep the engines running for a longer amount of time than the airplane will be flying. But this amount can be affected by strong winds and

changes in air pressure, which must be predicted and taken into account in the formulas used to decide on an amount of fuel. Surprisingly, the maximum weight of an airplane on take off is a higher value than the maximum weight of the airplane during the landing sequence. When flying a 747 jumbo jet, for example, the maximum weight that the airplane can manage to get off of the ground is about 870,000 pounds (394,625 kg). But the maximum weight of the aircraft that will enable a safe landing is about 630,000 pounds (285,793 kg). All loss of weight is due to the burning of fuel, and it is essential that enough fuel is used during the flight to bring the weight of the airplane down below the safe landing weight. Therefore, unless an airplane will be traveling the longest distance possible, the fuel tanks are rarely filled to their maximum capacity. In an emergency landing, the pilot must usually dump some of the fuel from the aircraft in order to lower the weight below the maximum safe landing weight. All of the factors that determine how much fuel to load into an airplane are calculated and rechecked using standard mathematical formulas that require a solid understanding of algebra.

SKYDIVING

In addition to a good deal of courage, the act of jumping out of an airplane involves a lot of algebra. In addition to the important calculations involved in any flight of an airplane, algebra is used by all skydivers to ensure that the plummet to Earth is as controlled as possible. For example, careful calculations are performed and rechecked in order to ensure that proper of size parachute is packed according to each diver's body weight.

Algebra is also needed to analyze the speed and acceleration of a diver. In turn, this analysis is critical to calculating the amount of time that a diver should wait to deploy the parachute after jumping out of the airplane. In a typical skydiving session, the pilot takes the airplane to an altitude of about 10,000 feet (3,048 m). After jumping, the average diver accelerates to a top speed (known as terminal velocity) of about 120 miles per hour (193 km/h). This gives a diver about 45 seconds of free fall (falling at terminal velocity), at which time the diver will be approximately 2,500 feet (762 m) above the ground. At this height, the diver must deploy a small parachute, called a drogue chute. The drogue chute is attached to the main parachute, and the main parachute is held in its container until the diver pulls a cord that allows the drogue chute to pull the main parachute out.

For the first jump, a diver is usually strapped to an instructor who makes sure that everything goes smoothly. This is known as a tandem jump, and requires different calculations to plan the jump safely. The small drogue chute is deployed almost immediately after exiting the airplane in order to slow the pair of divers for the duration of the free fall. If the drogue shoot were not open, the extra weight would cause the two divers to accelerate to a terminal velocity of up to 200 miles per hour (322 km/h), making tandem jumps inconsistent and unsafe. The main parachute remains in its container until the correct altitude is reached and one of the two divers pulls the release cord, allowing the drogue to open the main parachute. In a tandem jump, the main parachute must be much larger than the main parachute in a solo jump in order to stabilize the two bodies and slow them to a safe landing speed.

In another type of dive, called a high-altitude, low-open (HALO) jump, the diver jumps from an airplane traveling at a much higher altitude (often about 30,000 feet [9,144 m]) and does not open a parachute until reaching a significantly lower altitude than in a typical jump. In any jump higher than about 15,000 feet (4,572 m), divers must where oxygen masks because the air becomes too thin at higher altitudes. In a HALO jump, free fall can last for up to three minutes and the diver can reach speeds of over 200 miles per hour (322 km/h). This type of jump requires more training and preparation. Algebra provides the essential tools for performing the many calculations required to plan all of these different types of jumps.

Several algebraic formulas are used to analyze the effects that changes in the materials and shape of the main parachute have on a dive. Most parachutes are made of materials that allow no air to pass through them, making the parachute more effective for slowing the fall of the diver. However, if the parachute opens too quickly, it can slow the diver too quickly, possibly causing serious physical injury or damage to the parachute and other gear. To prevent the parachute from opening all at once, a mechanism is attached to the cords that hold the parachute to the diver. This mechanism slides slowly down the cords and controls the speed at which the cords can separate from each other, and in turn controlling how quickly the parachute opens after it is deployed. Algebraic formulas are essential for finding a safe speed at which the parachute should open, and for properly manufacturing the device that controls this speed.

Many older parachutes (and some still made for special purposes) are round, a shape that allows the diver to fall straight down in the absence of wind. Standard contemporary parachutes are rectangular in shape. These rectangular parachutes cause the diver to move forward while falling. The main benefit of a rectangular parachute is that it allows for much more directional control while falling, enabling smoother and more accurate landings.

To ensure safe, reliable operation, the dimensions of a parachute must be as close to perfect as possible. Luckily the specifications of all parachutes are determined and tested according to in-depth mathematical models. While these models are rather advanced applications of mathematics and physics, they rely heavily on algebraic reasoning.

Most divers employ an automatic activation device (AAD), a small computation device that performs constant calculations in order to deploy a reserve parachute if something goes wrong. The AAD unit is turned on when the diver is on the ground, and from then on it constantly monitors the altitude of the diver. When the diver jumps out of the airplane, the AAD senses that it is falling quickly, and is programmed to recognize this as the beginning of the fall. If the diver falls past a certain altitude without deploying the main parachute, the AAD shoots a piece of metal into the cord that holds the reserve parachute in place, deploying it automatically. As long as the reserve parachute opens correctly, the AAD will most likely save the life of a diver that is distracted or has lost consciousness. To make things more complicated, the AAD must also be programmed to differentiate a loss in altitude during free fall from a loss in altitude due to other events, such as the plane landing before the diver ever jumps out. This ensures that the AAD will only activate the reserve parachute if the diver is free falling.

Every aspect of skydiving—from the altitude and timing to the lengths of all the cords and the computer assistance of the AAD—involves the addition, subtraction, multiplication, and division of terms and expressions that represent many factors. An enormous amount of algebraic formulas helps divers, instructors, pilots, and equipment manufacturers understand the multitude of factors that must be controlled in every skydiving session.

CRASH TESTS

Every year, hundreds of new model cars, trucks, vans, and sports utility vehicles (SUVs) are purposely involved in controlled crashes in order to analyze the safety features of each model of automobile. The various components of these crash tests involve a seemingly endless amount of calculations. The slightest miscalculation can result in the recall of an entire model (which costs the manufacturer a substantial amount of money), and much worse, injury or death of people involved in real crashes. Therefore, the calculations involved in crash tests are checked multiple times under various conditions.

The design of crash test dummies, the main focus of all crash tests, involves a great deal of algebraic calculations. To ensure consistent results, all official crash tests use the same type of crash test dummy, belonging to the Hybrid III family of dummies. Various Hybrid III dummies are used to simulate different ages and body types for both genders. For each dummy, characteristics including height and weight are measured and factored in to the mathematical formulas used to analyze the amount of damage done to the dummy during an accident.

Crash test dummies must possess rather complex structures in order to simulate all of the parts of a human body that are usually affected in car crashes. For example, an elaborate spine consisting of metal discs connected by rubber cushions is attached to sensors that collect data used to analyze the damage done to the simulated spine. Sensors for measuring how quickly different parts of the body speed up and slow down are present throughout the body of a dummy. These sensors collect data that help to analyze the potential injury sustained due to the sudden decrease in speed caused by a crash (e.g., whiplash). Other sensors are placed throughout the dummy to measure the amount of impact endured by body parts (e.g., how hard the dummy's arm slams into the dashboard). Different colors of paint are applied to a dummy's various body parts so that, when an impact is detected by these sensors during in a crash test, researchers can determine which parts of the body collided with which parts of the car or airbag. A single sensor in a dummy's chest measures how much the chest is compressed due to the forces applied by the seatbelt and airbag.

All of the information collected by these sensors is injected into mathematical formulas in order to test and improve the timing and power of the seatbelts and airbags. For example, modern seatbelts sense abrupt decreases in an automobile's speed, immediately lock up but allow for a small amount of movement forward, then quickly increase the tension to bring the body to a stop, and finally decrease the tension so that the seatbelt does not cause injuries. In this way, the body slows down more gradually than it would if strapped in by a constantly stiff seatbelt. If the seatbelt stopped the body from moving forward all at once, the seatbelt itself could cause substantial injuries. Most cars now include airbags to supplement seatbelts in absorbing the forward force of the body and keeping it from slamming into anything solid. In order to effectively supplement a seatbelt, an airbag must deploy with perfect timing immediately after the seatbelt begins to lock up.

Algebraic operations are integral to the mathematical models used to analyze the various factors in a car crash. The wide range of problems solved using the mathematical models found in a crash test include the realistic design of dummies and the analysis of data collected

from their sensors; determining the effects of modifying the weight of the automobile and the load it carries; selection of the speed at which to hurl the automobile toward a concrete wall (frontal impact tests), or how fast to slam an object into the side of the automobile (side impact tests); analysis of the deliberate crunching of the materials that make up the automobile, which helps to absorb much of the impact; and calculation of the odds of surviving such a crash in real life. The rating systems used to indicate the effectiveness of an automobile's safety features also rely on algebraic formulas.

FUNDRAISING

In any fundraiser, the planners must be sure that enough money is brought in to cover the costs of the event and meet their fundraising goals. For example, to raise money for new equipment at a local hospital, the hospital's fundraising committee decides to sell raffle tickets for a new $30,000 car. The committee needs to raise at least $20,000 to be able to pay for the new equipment. To ensure that at least this amount of money is available after the paying for the car and the various components of the fundraiser, the committee decides to set up a mathematical model. To analyze the financial details of the event and decide the price of the raffle tickets and the minimum number of tickets that need to be sold, the committee prepares an algebraic formula. The formula will take into account the price of the car, the number of tickets sold, the price to be charged for each ticket, and the ultimate financial goal of raising $20,000. The formula they come up with is $G = TP - C - E$, where the variable G represents the amount of money to be raised, T is the number of tickets sold, P is the price of each ticket, C is the cost of the car, and E is the cost of the event. This equation states that the amount of money raised will be equal to the number of tickets sold multiplied by the price of each ticket, minus the cost, the car, and the event itself.

Because the purpose of this equation is to determine how many tickets to sell and at what price, the committee rewrites the equation with the term TP (representing the number of tickets multiplied by the price of each ticket) alone on the left side of the equation. By subtracting TP from both sides, the equation becomes $G - TP = -C - E$. Next, subtracting G from both sides gives $-TP = -G - C - E$. The term TP is now alone on the left side of the equation, but notice that all of the terms are negative. By multiplying both sides of the equation by -1, all of the negative terms become positive to yield $TP = G + C + E$. This equation now states that the amount of money that will be taken in from the sales of the raffle tickets is equal to the financial goal of the event plus the cost of the car plus the cost of the event itself. This equation allows the committee to substitute the values for the financial goal and the costs of the car and the event in order to determine the number of tickets that need to be sold, and at what price. However, the committee does not necessarily want the money brought in from the ticket sales to be exactly equal to the costs and fundraising goals; the money brought in needs to be greater than or equal to the money spent. Thus the committee makes this equation into an algebraic inequality by replacing the equal sign with the greater than or equal to symbol to get $TP \geq G + C + E$.

Next, the committee begins to plug numbers into their algebraic inequality. The committee's financial goal for the event is to raise $20,000, so $G = 20{,}000$. The car costs $30,000, so $C = 30{,}000$. The costs of the eventflyers, food, musical entertainment, renting a venue, and so onare estimated at $5,000, so $E = 5{,}000$. By substituting these values into the equation $TP \geq G + C + E$, the committee finds that $TP \geq 20{,}000 + 30{,}000 + 5{,}000 = 55{,}000$. Since $TP \geq 55{,}000$, the committee knows that the sale of tickets must amount to at least $55,000.

At a similar fundraising event in the previous year, a little over 12,000 raffle tickets were sold. To be safe, the committee decides to predict an underestimate of 11,000 raffle tickets sold at this year's event. Substituting 11,000 for the variable T, the inequality becomes $11{,}000P \geq 55{,}000$. Dividing through by 11,000 gives $P \geq 5$; so the committee needs to charge at least $5.00 for each ticket in order to pay for the car and the event, and have enough left over to pay for the new equipment. Since 11,000 was an underestimate for the number of tickets sold, the committee decides that it is safe to charge $5.00 for the tickets.

BUILDING SKYSCRAPERS

A massive amount of calculations are involved in all phases of skyscraper construction, from determining the amounts of time, manpower, money, concrete, steel, wiring, pipes, and paint needed to build the skyscraper, to determining the number of exits, bathrooms, and electrical outlets needed to serve the maximum capacity of people in the building. The calculations used in the actual creation of the structure are dependent on basic algebra; but long before construction can begin, more sophisticated mathematical formulas are developed to design a building that meets strict safety guidelines.

A skyscraper towering high above a city is susceptible to many unpredictable forces and must be able to withstand a wide range of punishing forces. These forces include large changes in weight due to people coming and going and precipitation collecting on the outside of the

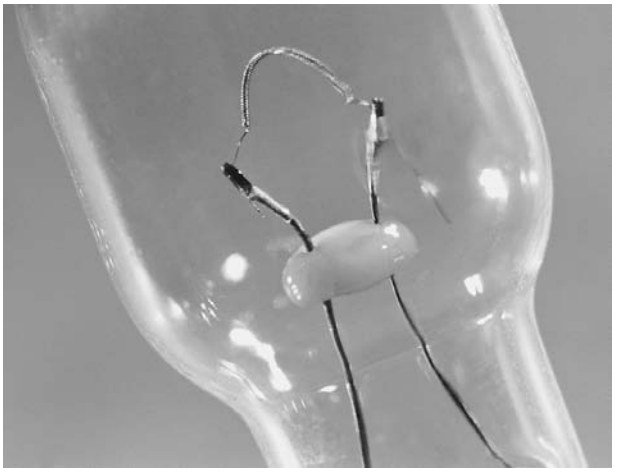


Figure 2: Light bulb filament. ROYALTY-FREE/CORBIS.

building, fluctuations in air pressure and wind, and seismic activity (earthquakes). A skyscraper must even be able to endure a sizeable fire or other direct damage to the structure of the building. For example, in 1945, a United States Army B-25 bomber, whose pilot had been disoriented by dense fog, crashed into the side of the Empire State Building, tearing gigantic holes in the walls and support beams, and igniting fires on five floors. However, the nearly 1,500-foot (457 m) tall skyscraper (the tallest in the world at the time) stood and the damage was repaired. If even small miscalculations had taken place in the planning of the Empire State Building, the crash might have caused the entire building to topple.

In-depth architectural specifications used to make a building visually pleasing and functionally efficient require countless algebraic systems. Formulas for modeling the various safety issues involved in constructing such a tall building take into account all of the factors that can compromise the structure of a skyscraper. The biggest problem to overcome when attempting to design a safe skyscraper is to make the structure stable enough to withstand wind and other forces. A skyscraper cannot be perfectly rigid. The structure must be allowed to sway slightly in all directions or its own weight would cause the structure to snap like a dry stick when acted on by forces like wind and earthquakes. On the other hand, if the skyscraper were allowed to sway too far from perfectly vertical, the building would fall over. Under normal conditions, the movement of a skyscraper is undetectable by the human eye, and unnoticed by occupants. The amount of flexibility in the structure must be controlled perfectly by the structure of each floor. Modeling the nature of a skyscraper's flexible components involves the use of the imaginary number, i, where $i^2 = -1$.

In the study of basic algebra, the value of i is not logical because multiplying any real number by itself results in a positive number, e.g., $2^2 = (-2)^2 = 4$. Multiples of i, such as $2i$ and $-3i$, are called imaginary numbers or complex numbers. An entire field of mathematics, known as complex analysis, is devoted to the study of the properties of imaginary numbers. Although imaginary numbers do not follow the rules of basic algebra, they are often used to simplify enormous, intricate polynomial equations—like those used to model the stability of skyscrapers—into more manageable equations. After an equation is solved using imaginary numbers, the solution can often be transformed back into real numbers. The use of imaginary numbers enables mathematicians and scientists to solve problems that would otherwise be unsolvable. For example, by assuming that i exists and using it in algebraic expressions, mathematicians, physicists, chemists, statisticians, and engineers are able to model and simplify complicated phenomena. In addition to modeling the slight swaying of a skyscraper, imaginary numbers can be used to model the behavior of electrical circuits, the springs that absorb shock in automobiles, and sophisticated economic systems.

BUYING LIGHT BULBS

Incandescent light bulbs produce light by passing electricity through a thin metal coil, called a filament (see Figure 2). When electricity is passed through the filament, it glows and illuminates the light bulb. The electricity also produces heat as it passes through the filament. In fact, special bulbs, called heat lamps, are intended to produce heat for purposes such as heating food and drying plants; but in most light bulbs, heat is an undesired (and unavoidable) side effect which eventually causes the filament to burn out. The amount of time that a light bulb can be turned on for before it is expected to burn out is printed on most packages so that shoppers can compare the life expectancy of the various available bulbs. It turns out that the amount of electricity that passes through the filament when the light bulb is turned on is all that is needed to predict how long a light bulb will last.

By logging and analyzing the results of various experiments to test the life of light bulbs under different conditions, the life expectancy of a light bulb has been found to be inversely proportional to the voltage that is applied to the filament. That is, the life expectancy is equal to some number divided by the number of volts raised to a power; when two values are inversely proportional, decreasing one value causes an increase in the other value. The life expectancy of most incandescent

light bulbs is inversely proportional to the 12th power of the applied voltage and can be expressed as $L = a / V^{12}$, where L represents the life expectancy, a is a constant, and V represents the applied voltage. This expression indicates that using a lower value for the variable V results in the constant a being divided by a smaller number as long as V is >1, so that the life expectancy L is equal to a larger number. This means that lowering the voltage that is allowed to surge through the filament increases the life expectancy of the light bulb.

On the other hand, less light is produced if less electricity is allowed to pass through the filament. The amount of light that is produced is dependent on the voltage and can be expressed as $X = bV^{3.3}$ where X represents the amount of light produced by the light bulb, b is a constant, and V (which is raised to a power of 3.3) represents the voltage. In contrast to the relationship between voltage and life expectancy, it is said that the amount of light produced is directly proportional to the voltage, meaning that the amount of light is equal to some number multiplied by the voltage raised to a power. In this type of relationship, increasing the value of one variable also increases the value of the other variable. As can be deduced by examining these two equations, lowering the voltage has a much smaller effect on the amount of light produced than it does on the life expectancy of the light bulb. In other words, a small decrease in voltage increases the life expectancy by a relatively substantial amount, but decreases the amount of light produced only slightly. Therefore, buying light bulbs with lower voltage values usually increases the amount of time before light bulbs need to be replaced without resulting in a marked decrease in illumination.

ART

Though art is often seen as an pure expression of creativity, artists cannot help but use mathematics in the creation of any piece, whether the artist realizes it or not. The size and dimensions of a canvas and frame are carefully chosen by a painter; and though the choice is often dependent only on the artist's preference, the measurements of the rectangular canvas can be analyzed to determine which will be the most pleasing to the average person, or which will help to effect the emotions that the

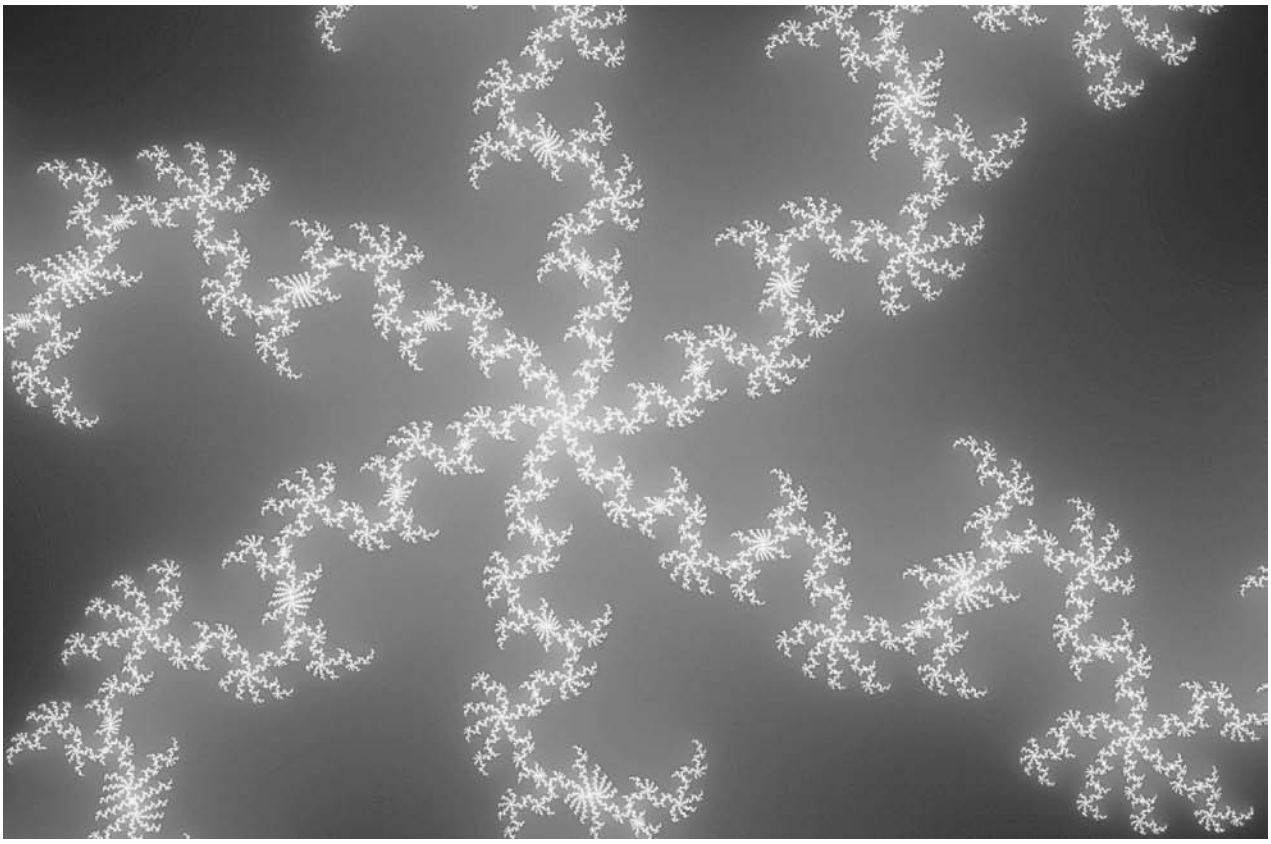


Figure 3: Fractals in art. BILL ROSS/CORBIS.

artist hopes convey in the painting. The colors of the paints can also be examined to determine the best mixtures of the primary colors, or to analyze the emotions effected by different pigments and patterns.

In the age of computers, art has expanded its definition to encapsulate computer-generated art, including intricate and realistic images, and fractal images automatically created by computer programs using mathematical formulas (see Figure 3). Fractals are actually complex geometric shapes; but just as algebra can be used to define and analyze the characteristics of a circle or rectangle, algebraic sequences can be used to create and investigate fractal images. In a fractal, each part of the pattern has the same characteristics as every larger part. That is, when part of the fractal is magnified, it is generally undistinguishable from the original image. In many fractal images, a large part has small copies of itself sticking out of it, and these copies have small copies of themselves sticking out of them. For example, a triangle appears to have breaks in it, and smaller triangles fill the gaps so that the line is continuous. These smaller triangles have breaks in them, and even smaller copies of the triangle fill those gaps. Theoretically, this pattern can repeat infinitely so that no matter how many times the image is magnified, the same pattern will appear.

These infinite patterns are defined by special algebraic constructions, known as infinite series, which are determined by infinitely repeating numerical patterns. An infinite series repeats infinitely with, for example, each term raised to the power equal to the number of terms that precede the term in the series. Such an infinite series defines a fractal image similar to a snowflake, with the resulting image like a six-sided star, where each point can be thought of as a triangle missing one of its sides. The sides of each triangular point are cut at regular intervals and filled in with the next smaller size of triangles in the pattern. As the series continues, smaller and smaller triangles are added to the image. In theory, the series continues on forever and the image contains infinitely smaller and smaller triangles.

Though fractal images are frequently used to create beautiful artistic graphics, they have applications in other computer imaging projects as well. For example, computer generated maps use fractals to create realistic coastlines and mountain regions. No matter the use, fractal images are created by defining infinite series that involve algebra at every step.

POPULATION DYNAMICS

In any population—including bacteria, ants, fish, birds, and humans—many factors contribute to the number of individuals present at a given time and the rate at which the population increases or decreases. Some of the most common and important factors include the availability of food, the abundance of predators, and the inherent reproductive capacity and natural mortality rate of the species. In most investigations of population dynamics, researchers attempt to set up algebraic formulas using terms that represent all of the pertinent factors that determine the way that a population fluctuates. Basic population models are similar to $N = aZ + bY + cX + dW + eV + fU$, where N is the current number of individuals in the population. Each term on the other side of the equation represents a different factor in the life of the species. The variables Z, Y, X, W, V, and U represent different factors. These variables often take into account the number of individuals in the population in the immediate past (e.g., reproductive rates of are dependent on the number mature individuals). The constants a, b, c, d, e, and f are the coefficients of the terms and define the extent to which each represented factor affects the population. A large coefficient indicates that the term has a relatively significant effect, while a coefficient smaller than one indicates that the term has a relatively minimal effect. The coefficients for factors that decrease the number of living individuals (e.g., mortality rate and abundance of predators) have negative values, essentially subtracting individuals from the total. Positive coefficients are attached to terms representing factors such as reproductive rate and availability of food. This type of algebraic formula has helped to save many endangered species by facilitating important research of the affects that human developments have on wildlife populations around the world.

FINGERPRINT SCANNERS

Security is an essential consideration for many organizations, including police and military groups, financial institutions protecting money, and hospitals protecting sensitive medical records. All forms of security can be penetrated by an experienced attacker. Although many security devices attempt to give the appearance of being impenetrable, the true goal of a solid security system is to minimize the number of successful attacks by maximizing the time and energy required to circumvent the implemented security measures. The odds of an attacker successfully breaking a security system can be calculated using algebraic expressions that represent the various factors involved (e.g., the thickness of a safe door, or the odds of guessing a given password).

A form of identification steadily increasing in popularity, called biometrics, involves comparing an individual's unique physical characteristics with previously

stored data about the individual. When someone first uses a biometric device, physical characteristics are measured, translated into mathematical formulas by a computing device, and stored for future comparison. The most widely developed biometric security devices include cornea and iris scanners that measure the characteristics of the parts of an individual's eyes; face scanners that can recognize major facial features; voice scanners that measure the frequencies in an individual's voice; and fingerprint scanners that read and interpret the unique curves and patterns of an individual's fingerprints.

Fingerprint scanners are widely accepted as one of the most effective forms of identification, and are becoming more common in all types of secure environments. The uses of fingerprint scanners range from the physical protection of a secured room to the protection of sensitive computer files. Many computer mice and keyboard manufacturers integrate fingerprint scanners into their products in hopes of replacing passwords as the most common form of identity verification for personal computers. An increasing number of automobile manufacturers have begun to incorporate fingerprint scanners into door locking mechanisms and ignition systems, so that the owner of a vehicle does not need a key to lock and unlock the car, or start the engine. Banking institutions are also beginning to look to fingerprint technology in hopes of replacing bank cards and personal identification numbers (PINs).

Like all biometric devices, fingerprint scanners map the unique characteristics of a fingerprint into mathematical formulas, which are used later to determine whether or not the fingerprint present on the scanner matches stored data. The size and relative location of the prominent features in each fingerprint are represented by the terms of mathematical formulas, so fingerprint scanners utilize massive amounts of algebraic operations during each security session.

An bar-coded identification bracelet is scanned at Georgetown University Hospital in Washington, D.C. A/P WIDE WORLD. REPRODUCED BY PERMISSION.

Potential Applications

TELEPORTATION

Throughout history, humans have invented increasingly advanced methods of transportation—from the invention of the wheel to the first trip into space—in order to enable and expedite the process of traveling from one physical location to another. However, even with all of the advances in transportation, no vehicle can take passengers from one point to another without traveling across the space in between. Learning how to skip the intermediate locations is the goal of scientists who are attempting to invent and perfect teleportation devices. Similar to the fantastic idea first popularized in science-fiction, teleportation devices essentially collect information about an object, destroy the object, and send the information about the object to a different location, where the object is reconstructed. In a sense, this idea is similar to a fax machine, which translates a copy of a document into numerical information, and sends the information to another fax machine that uses the information to construct and print another copy. A teleportation device collects information about an object and translates it into numerical information according to mathematical formulas. Different formulas are then used to reconstruct the object elsewhere.

In 1998, a group of physicists performed the first successful teleportation experiment. In the experiment, information about a photon (a particle of energy that carries light) was collected, sent through a cable one meter in length, and used to construct a replica of the photon. When the replica was created, the original photon no longer existed, so the photon is considered to have traveled instantaneously to a different location.

It is uncertain whether or not this type of travel will ever be safe for living organisms. The idea of being destroyed and reconstructed elsewhere sounds rather foreign and frightening, and it will surely be difficult to find willing subjects for experiments. However, the ability to teleport energy will likely have profound effects on computer networks. Instead of using cables or airwaves to transfer information between computers, information will be instantaneously teleported from machine to machine, essentially eliminating delays in the transfer of information.

Figure 4: SpaceShipOne: the future of travel? JIM SUGAR/CORBIS.

PRIVATE SPACE TRAVEL

A new form of transportation will most likely revolutionize the way that humans think about space travel. SpaceShipOne (see Figure 4) is the first manned spacecraft project that does not depend on government funds. This privately owned and operated craft is intended to take anyone who can afford a ticket on a brief trip into space. In order to alleviate the most difficult part of any flight into space—the launch from the ground—SpaceShipOne is launched from a second aircraft, called White Night. While attached to White Night and during the launch into space, SpaceShipOne is in a contracted configuration. After the spacecraft launches and reaches its highest altitude, it spreads its wings in a configuration that slows its decent back into the inner atmosphere. Finally, the craft is reconfigured again to work much like an airplane, allowing the pilot to safely steer and land.

Because this project is not funded by the government, the company that designed and built SpaceShipOne and White Night had to create their innovative technology from scratch, a task that involves an unimaginable amount of calculations, all of which rely on the fundamental concepts of algebra. Developing a rocket propulsion system alone involves a multitude of mathematical formulas. Designing the three different configurations of SpaceShipOne also involves an enormous amount of mathematical models for determining the optimal size and shape of the various parts of the spacecraft. The idea of an average private citizen having the opportunity to travel into space on a regular basis is a shining example of the endless potential for using algebra to explore the real world.

Where to Learn More

Books

Johnson, Mildred. *How to Solve Word Problems in Algebra*, 2nd Ed. New York, NY: McGraw-Hill, 1999.

Ross, Debra Anne. *Master Math: Algebra*. Franklin Lakes, NJ: Career Press, 1996.

Periodicals

Backaitis, S.H., H.J. Mertz, "Hybrid III: The First Human-Like Crash Test Dummy." *Society of Automotive Engineers, International*. Vol. PT-44 (1994): 487–494.

Key Terms

Algebra: A collection of rules: rules for translating words into the symbolic notation of mathematics, rules for formulating mathematical statements using symbolic notation, and rules for rewriting mathematical statements in a manner that leaves their truth unchanged.

Arithmetic: The study of the basic mathematical operations performed on numbers.

Coefficient: A coefficient is any part of a term, except the whole, where term means an adding of an algebraic expression (taking addition to include subtraction as is usually done in algebra. Most commonly, however, the word coefficient refers to what is, strictly speaking, the numerical coefficient. Thus, the numerical coefficients of the expression $5xy^2 - 3x + 2y -$ are considered to be 5, -3, and $+2$. In many formulas, especially in statistics, certain numbers are considered coefficients, such as correlation coefficients.

Constant: A value that does not change.

Equation: A mathematical statement involving an equal sign.

Exponent: Also referred to as a power, a symbol written above and to the right of a quantity to indicate how many times the quantity is multiplied by itself.

Formula: A general fact, rule, or principle expressed using mathematical symbols.

Term: A number, variable, or product of numbers and variables, separated in an equation by the signs of addition and equality.

Variable: A symbol representing a quantity that may assume any value within a predefined range.

Web sites

How Stuff Works. "How Skyscrapers Work." Science Engineering Department. <http://science.howstuffworks.com/skyscraper.htm> (March 22, 2005).

NASA. "Beginner's Guide to Aerodynamics." Glenn Research Center. March 4, 2004. <http://www.grc.nasa.gov/WWW/K-12/airplane/bga.html> (May 27, 2005).

Algorithms

Overview

An algorithm is a set of instructions that indicate a method for accomplishing a task.

Algorithms—often used subconsciously—help us solve a wide range of problems in everyday life. Algorithms can be written to describe the method to tie a shoelace, bake a cake, or address an envelope. Algorithms are composed of the steps needed to accomplish a task and are written in such a way that no "judgment"—other than the fact that a particular step has been performed—is required to accomplish the overall task.

Fundamental Mathematical Concepts and Terms

In mathematics, an algorithm is a method for solving a mathematical problem by using a finite number of computations that repeat certain operations or steps. Not all algorithms lead to a single solution (deterministic algorithms), some algorithms can be designed that lead to multiple solutions (nondeterministic algorithms).

The length of time required to complete an algorithm is directly dependent on the number of steps involved and the speed with which the steps are completed. For example, a young child might take several minutes to add a long column of numbers—an algorithmic task performed by most computers in a fraction of a second.

A Brief History of Discovery and Development

The term algorithm is derived from the name of the ninth century Arabic mathematician and Tashlent cleric al-Khowarizmi, the mathematician most often credited with the early development of algebra.

With the rise of an industrial mechanized society, algorithms were developed to control a broad array of devices and procedures from traffic signals to the operation of production lines. Algorithms were used in almost every facet of communication and control (e.g., in routing aircraft at designated flight levels and speeds).

Microchip technology has increased the computational speed of computers so that, by using algorithms, they can quickly scan large arrays of data. For example, computers can use algorithms based upon the rules of chess to quickly evaluate the outcome of potential chess moves. Although the human brain is far more complex than even the most powerful supercomputers, high-speed supercomputers

using well-designed algorithms have sometimes been able to defeat world chess champions in test matches.

Real-life Applications

OPERATIONAL ALGORITHMS

Probably the most commonly used algorithm is one used in the operation of addition. This algorithm is used everyday in countless ways, and is so basic to mathematics that most people do not realize that they are using an algorithm when adding numbers.

The addition algorithm relies upon Hindu-Arabic positional notation—the most commonly used notational system that imparts a value to the position of a numeral. Each position or column is 10 times larger than the column or position to its right (e.g., the number 3,456 is interpreted as the sum of 3 "thousands," 4 "hundreds," 5 "tens," and 6 "ones").

This positional notation and addition algorithm makes possible the easy addition of large numbers, and long columns of numbers. The addition algorithm, the repetitive steps used to add numbers, specifies that we count by "ones" in the right hand column, by "tens" in the next column to the left, by "hundreds" in the next column to the left and so on. When the sum in a column exceeds nine, the amount over 10 is retained and the rest is carried to the next column to the left. To add 67 and 97 we add each column. Adding each column gives us 14 "ones" and 15 "tens." Using the addition algorithm, the 14 "ones" are equal to 1 "ten" plus 4 "ones" and so we carry one ten to the column to the left. This then gives us 16 "tens" and 4 "ones." The additional algorithm dictates that 10 "tens" are equal to one hundred and so the number 1 is inserted into the "hundreds" column and the remainder of 6 left in the "tens" column. The algorithm thus yields the correct answer of 164 ($67 + 97 = 164$).

The addition algorithm does not work for other systems of notation, such as Roman numerals.

Various methods are used to teach the essentials of the addition algorithm, so the description above may not be exactly what you remember from early elementary school. Regardless, whatever the words you use to describe the operation of addition—and the other operations of subtraction, multiplication, and division—those terms describe an algorithm in action.

ARCHEOLOGY

Archeologists (scientists who study past civilizations) collect as much information as possible as they explore a site or find. A small grave or ancient trash dump may ultimately result in thousands of measurements of pieces of bone, pottery, or other fragments of the past. Other scientists who study the past, including archaeo-astronomers (scientists who study and make calculations about what past civilizations may have observed in the skies with regard to the movement of the Sun, planets, and stars), also compile thousands of observations to yield clues about ancient humankind. Algorithms are used to analyze those mountains of data to yield clues regarding the building of ancient temples, monuments, and tombs or upon the movements and cultural practices of ancient civilizations.

COMPUTER PROGRAMMING

Computers are particularly adept at utilizing algorithms, and algorithms lie at the heart of computer programming, the set of instructions that computers use to analyze data. The creation of elegant (a term used by mathematicians to describe something simple yet powerful) and thus faster algorithms has become an important consideration in the study of theoretical computer science.

Logical algorithms (rules and steps based upon patterns of mathematical logic or proof) including a class of algorithms known as "backtracking" algorithms were developed in the 1960s to explore methods of solving computational problems. Such algorithms can be designed to test possible combinations of sub-problems and such algorithms result in tree-like solutions. A particular solution can be traced back to through prior solutions that are analogous to backtracking through the increasing larger more inclusive branches of a tree that ultimately lead to the trunk. Navigating the solution tree, analogous to a squirrel climbing through the limbs of a tree, produces computational solutions that can then be described as "longest" or "shortest" path solutions. Many computer-programming languages rely on backtracking. For example, if a particular sub-problem solution (a particular branch of the solution tree) proves to be incorrect, the computational algorithm "backtracks" and tries another path to solve the problem.

CREDIT CARD FRAUD DETECTION

Every time you use a credit card, the purchase made is analyzed by computers programmed with algorithms to detect the crime of credit card fraud. Banks and financial institutions that issue credit cards program their computers to use a series of algorithms that compare each purchase to the established pattern of use. For example, if you usually use your credit card in the New York area to

make purchases totaling $150 a month, the computer's algorithms should be able to quickly determine that a series of charges from a foreign country totaling thousands of dollars is "unusual." The fact that the purchase does not fit an established pattern can be determined by algorithms and is often enough to alert bank security officials that further investigation is required before authorizing a particular purchase. Upon investigation, they may discover that the user is the authorized cardholder enjoying a vacation or semester abroad. On the other hand, investigation may reveal that the credit card number has been stolen, and that the intended purchase is unauthorized. The use of algorithms to analyze purchases can thus save the bank—or credit card holder—thousands of dollars.

CRYPTOLOGY

In 1977, Ronald Rivest, Adi Shamir, and Leonard Adleman published an algorithm (known as the RSA algorithm—a name derived from the first letters of the founder's last names) that marked a major advancement in cryptology. The RSA algorithm factors very large composite numbers. As of 2004, the RSA algorithm was the most commonly used encryption and authentication algorithm in the world. The RSA algorithm was used in the development of Internet web browsers, spreadsheets, data analysis, email, and word processing programs.

DATA MINING

Association of data in a data mining process involves the use of algorithms that establish relationships or patterns in data. Such algorithms use "nested" or sub-algorithms that rely on statistics and statistical analysis to make associations between data. Usually the algorithm designer (e.g., a computer programmer) specifies desired associations or patterns to be established. Algorithms can be written, however, to perform what is termed exploratory analysis, a form of analysis where associations between data are sought without a preconception or guess as to what patterns might exist.

DIGITAL ANIMATION AND DIGITAL MODEL CREATION

Moviemakers rely on mathematical algorithms to construct digital animation and models. Such algorithms relate points on known surfaces to points on a drawing (often a computer drawing) or digital model. For example, data points for the movement of an arm or leg can be obtained by actors wearing special gloves or sensors that translate movements such as waving or walking into data (sets of numbers) that can be analyzed by algorithms designed to fill in the gaps between data points. Such algorithms allow animation experts to subsequently draw and animate figures with increasingly realistic features and movement. Model makers can construct digital models at a fraction of the cost needed to construct and test physical models.

DNA OR GENETIC ANALYSIS

Biochemists use algorithms, more commonly referred to in the laboratory as "lab procedures" to identify DNA markers that allow scientists and physicians to determine genetic relatedness (identification of parents or family members) to settle a court case or find a suitable organ donor, determine a patient's risk of disease susceptibility risk, or to evaluate the effectiveness (efficacy) of drug treatments.

In addition to physical testing algorithms, mathematical and computer algorithms can be used to determine or predict patterns of genetic inheritance. The pundit square used in beginning biology is a simple yet powerful use of algorithms that result in the diagrammatic representation of potential gene combinations. In some cases, the pundit square allows the calculation of the odds of having a child who might develop sickle cell anemia or be a carrier of the gene that might lead to actual sickle cell disease in their children.

The task of analyzing massive amounts of data generated by DNA testing is daunting even for very powerful computers. New technologies, including so-called "bio-flip" technologies use specialized computer algorithms to detect small and differences and changes in the structure of DNA (i.e., variation in genetic structure).

ENCRYPTION AND ENCRYPTION DEVICES

Although the technology exists to allow the construction of cryptographic devices intended to protect private communications from unauthorized users while at the same time assuring that authorized government agencies (e.g., those agencies such as the FBI who might obtain a court order) can quickly decode (decrypt) and read messages as needed, such technologies remain highly controversial. So-called "clipper-chips" and "capstone chips" would allow use United States law and intelligence agencies to use specific algorithms to decode encrypted (coded) messages. Certain authorized agencies would then hold the algorithmic "keys" (the step-by step procedures and codes) to any communication using the encrypting technology.

Use of the clipper chip was first adopted and authorized in 1994 by the National Institute of Standards and Technology (NIST). The United States Department of the Treasury was initially designated to hold the keys (algorithms) to decode messages. Rules regarding access to the keys are defined in state and national security wiretap laws. The clipper chip utilizes the SKIPJACK algorithm—a symmetric cipher (code) with a fixed key length of 80 bits. A bit is shorthand for "binary digit," a unit of information (a "1" or "0" in binary notation).

A cipher uses algorithms (i.e., sets of fixed rules) to transform a legible message (clear text or plaintext) into an apparently random string of characters (ciphertext or coded text). For example, a cipher might be defined by the following rule: "For every letter of plaintext, substitute a two-digit number specifying the plaintext letter's position in the alphabet plus a constant (or key) between 1 and 73 that shall be agreed upon in advance." This would result in every letter in the alphabet being represented by a number between 17 and 99 (depending on the particular constant used). For example, if 16 is the agreed-upon constant, then the plaintext word PAPA enciphers to 32173217 as follows: $P = 16 + 16 = 32$; $A = 1 + 16 = 17$; $P = 16 + 16 = 32$; $A = 1 + 16 = 17$. Real keys would, of course be longer and more complex, but the basic idea remains the same: an algorithm-based encryption key allows messages to be locked (enciphered) or unlocked (deciphered), just as a physical key fits into a lock and allows it to be locked and unlocked. Without a key, a cipher algorithm is missing its most critical part. In fact, so important is the key that many times the algorithm itself is widely known and distributed—it is only the keys that remain secret. For this reason, the algorithm used to code messages may remain the same for months or years, but the keys change daily.

Other algorithms remain a mystery. In 1943, Alan Turing (1912–1954), Tommy Flowers (1905–1998), Harry Hinsley (1922–1998), and M. H. A. Newman at Bletchley Park, England, constructed a computational device called Colossus to crack the Nazi German encryption codes created by the top secret Enigma machine used by the Germans. The decryption algorithms used by Colossus remain secret.

In the late 1970s, the United States government set a specific cipher algorithm for standard use by all government departments. The digital encryption standard (DES) is a transposition-substitution algorithm that offers 2^{56} different possible keys (a number roughly equivalent to a 1 followed by 17 zeroes). As larger a number of different keys as that number represents, modern higher speed computers might allow hackers (who also use algorithms) to too easily crack codes with this many keys, and so a new algorithm, known as the advanced encryption standard, is replacing the old algorithm.

Security is increasing as a function of who can develop the most sophisticated algorithms to either protect data, or hack into protected algorithmic codes.

IMAGING

The digitization of images used in modern digital computers would not be possible without the use of algorithms to translate the images into numbers and back again into a viewable image. Digital cameras can be in a vacationer's backpack or be mounted in satellites in orbit hundreds of miles above the Earth. Digital cameras offer higher resolution (the ability to distinguish small objects) than cameras that use light-sensitive photographic film. Digital photo manipulation has also revolutionized photography, including commercial advertising, and offers new security challenges to uncover altered photos. Fractal image compression algorithms allow much greater compression in the storage of images.

Digital cameras capture reflected light on a chip or charged coupled device (CCD). The surface of the CCD contains light-sensitive cells (photo diodes). Each cell or diode represents a pixel and so the pixel becomes a basic unit of a digital image. Light-stimulated diodes produce a signal (often using a transistor) with a voltage that corresponds to the light intensity recorded by the diode. An algorithm in the camera's processing unit then translates that signal into binary code—1s and 0s—that can later be reconverted by the reverse algorithm back into a viewable image. For example, algorithms may assign a code sequence between 0 and 255 to color data (0 is black and 255 indicates an intense red color). These codes are then turned into eight digit binary code sequences (00000000 for black, 111111111 for the most intense red).

Digital photo manipulation involves the alteration of the binary code (i.e., the digital 1s and 0s) that represents the image. While algorithms can be used to alter photos, they can also be used to detect forgery or alteration. Algorithms can compare values of pixels in the background of an image and determine whether they are consistent. Other parts of an image can be protected by altering certain pixels to form a digital watermark that can only be removed by application of a particular algorithm to the image binary code.

INTERNET DATA TRANSMISSION

All information sent by a computer over the Internet contains the sending computer's hardware source address

(MAC address). This is similar to a return address included on a piece of physical mail (snail mail). Conversely, all the information that the computer accepts must be addressed to its unique hardware address (or often a more common "broadcast address" that is similar to a zip code used in regular mail). When an packet of data (e.g., a portion of a text) is received, the receiving computer subjects the incoming packet of data to a processing algorithm, a mathematical formula or set of procedures that determines whether the address information is correct and the message intended for that computer. If the packet of data is accepted, additional algorithms are used to decode the binary message (a series of ones and zeros such as "100010110" into text, pictures, or sound).

MAPPING

Algorithms can analyze data measurements of height, depth, and distance to construct maps. For example, bathymetric maps (maps that depict the oceans as a function of depth) help develop a model of a body of water as depth increases. Such maps are important to fishermen and similar algorithmic programs analyze data in navigational and "fish-finding" equipment aboard many commercial and sport fishing boats.

To construct a precise map of the region, whether of land or at sea, it is necessary to perform detailed measurements, a task increasingly performed by satellites (or in the case of bathymetric maps, by ships with echo sounding surveying equipment that bounce sound waves off the ocean floor). Such data is then set into an array (a particular grid or pattern) that are analogous to strips of a map. Algorithms perform calculations that link the data between various strips and allow the construction of a larger map of the area.

THE GENETIC CODE

Humans themselves are the result of an molecular algorithm that operates at the genetic and cellular level. The genetic code ultimately relates a sequence of chemicals called nitrogenous bases found in deoxyribonucleic acid (DNA) to the amino acid sequences of proteins (polypeptides). These proteins control the biochemistry of the body. The algorithm that describes this process allows scientists to understand the genetic and molecular basis of heredity and many genetic disorders.

In humans, DNA is copied to make mRNA (messenger RNA), and mRNA is used as the template to make proteins. Formation of RNA is called transcription and formation of protein is called translation. This process is the fundamental control mechanism for the development (morphogenesis), growth and regulation of the body and complex physiological processes.

The structure of DNA—and the sequences formed during transcription—can, for example be predicted from an algorithm (based upon the physical shape of the molecules themselves) that specifies that the nucleotide with the nitrogenous base adenine will pair only with a nucleotide that contains a thymine base (an A-T pair). Likewise, nucleotides with a cytosine base will pair only with a nucleotide that contains a guanine base (a C-G pair). The molecular algorithm allows the prediction of bases (e.g., the ATTATCGG sequences) that in triplet sequences (three base sequences) then form the backbone of genes.

A sequence, such as A-T-T-C-G-C-T . . . etc., might direct a cell to make one kind of protein, while another sequence, such as G-C-T-C-T-C-G . . . etc., might code for a different kind of protein.

MARKET OR SALES ANALYSIS

Algorithms are routinely used to analyze buying and selling patterns. Businesses rely on algorithms to make decisions regarding which products to sell, or to which portion of the overall market advertising can be most effectively targeted (e.g., an advertising campaign designed to sell blue jeans to teenagers). Such marketing algorithms make associations between buying patterns and established demographic data (i.e., data about the age, sex, race, income groups, etc.) of the user.

For example, market specialists use algorithms to study data developed from test groups. If a product receives a sufficiently favorable rating from a very small test group, a manufacturer may skip the costs and delays of further testing and move straight into production of a product. The algorithms might be simple (e.g., if 75% of the initial test group likes the product the manufacture knows from experience that there will be sufficient sales to make a profit) or complex (e.g., a complex relation or weighting of responses to the demographics of the test group). Less enthusiastic results may require further testing or a decision not to develop a particular product.

Algorithms can also be used to compare observed responses of a test group to anticipated responses (or responses from other test groups) to determine which products might gain a market advantage if manufactured in a certain way. For example, algorithms can be used to analyze data to find the most favored color, size, and potential name of a skateboard and intended for sale to 10–13-year-old boys.

LINGUISTICS, THE STUDY OF LANGUAGE

The study of the elements of language (e.g., English, French, ancient Native American languages, etc.) is increasingly a quantitative science that relies on mathematical analysis to yield clues about the source of a language and the placement of a language within larger families of languages. This quantitative analysis often requires sophisticated computer-processing and algorithms designed to sift through large databases to determine statistically significant relationships (more than accidental or random relationships) between words and the use of words.

SECURITY DEVICES

Closed circuit television (CCTV) is increasing a part of security measures. CCTV is widely used in the United Kingdom (CCTV is so widespread in London, for example, that it can be used to detect violations of traffic laws) and its use is growing in Europe and in the United States. Many major cities, including New York and Washington, D.C., now utilize widespread public-surveillance CCTV systems, most often operated by the local police.

As a response to terrorism, images from CCTV cameras, especially those located in airports and other transportation hubs, are analyzed by algorithms that compare biometric data (e.g., height, shape of head, etc.). Facial recognition systems use algorithms to compare observed facial features that are hard to change (e.g., head size, width/length of nose, distance between the eyes) against databases containing photographs of known terrorists and other criminals.

SPORTS STANDING AND "SEEDINGS"

Many sports tournaments such as the NCAA men's and women's basketball championships—or the pairings for the annual football bowl games that are now used as "national championship" games—rely on algorithms that are designed to take the bias (an unwarranted predisposition in favor of someone or something) and prejudice (an unwarranted predisposition against someone or something) out of the selection process.

Depending on the sport, "seeding" algorithms can be designed to use data derived from poll results, strength of schedule points, points earned through competition, individual race or game results, conference or league standings, etc. Algorithms are used to determine everything from which lane a runner or swimmer starts a particular race to more fundamental questions as to whether an athlete or team is invited to participate in a competition.

Trevecca Nazarene's Alex Renfroe drives around Lewis-Clark State's Danny Allen during the second half of a 2005 NAIA Division I tournament game. Algorithms are often used in sports to determine bids and placement of teams in match brackets. AP/WIDE WORLD PHOTOS. REPRODUCED BY PERMISSION.

For example, the NCAA algorithms were designed to replace a simple reliance upon polls of coaches or sportswriters than were often driven by publicity and favoritism toward certain teams or teams from certain areas of the country. Although the new algorithm-driven selection processes are not perfect, they are an attempt to make the selection processes more fair.

TAX RETURNS

Some tasks that seem complex and difficult can be broken down into simpler steps (i.e., an algorithm can be written to accomplish the task). For example, completing a tax return form can be a time consuming and frustrating task for many people—especially young students who may have just started working.

At first glance, tax forms often seem overly and needlessly complex. The forms, however, can be completed by using a series of algorithms that are described in the instructions for each form. The instructions themselves are keyed to the step-by-step (systematic) completion of the tax form they describe. Completing a series of smaller,

Key Terms

Algorithm: A fixed set of mathematical steps used to solve a problem.

Operation: A method of combining the members of a set so the result is also a member of the set. Addition, subtraction, multiplication, and division of real numbers are everyday examples of mathematical operations.

Program: A sequence of instructions, written in a mathematical language, that accomplish a certain task.

usually less complex steps, allows taxpayer to correctly complete the tax form.

Potential Applications

ARTIFICIAL INTELLIGENCE

Since their development, electronic computers have been programmed with algorithms to accomplish specific tasks (e.g., a numerical calculations). Current research and development seeks to develop sophisticated algorithms that when combined with more flexible rules of operation may result in what is termed "artificial intelligence." The exact differences in computers that perform intricate algorithms and those with "artificial intelligence" is often a hotly debated topic among scientists and engineers. Regardless, one element or defining characteristic of artificial intelligence that is widely agreed upon is that a computer using artificial intelligence will use flexible rules rather that rigid algorithms for seeking solutions. Artificial intelligence programming may even allow computers to modify their own programming rules and develop their own algorithms for tackling problems.

Where to Learn More

Books

Edelstein, Herbert A. *Introduction to Data Mining and Knowledge Discovery*, 3rd ed. Potomac, MD: Two Crows Corporation, 2000.

Grahm, Alan. *Teach Yourself Basic Mathematics.* Chicago: McGraw-Hill Contemporary, 2001.

Sherman, Chris, and Gary Price. *The Invisible Web: Uncovering Information Sources Search Engines Can't See.* Medford, NJ: CyberAge Books, 2001.

Web sites

National Institute of Standards and Technology. "Advanced Encryption Standard: Questions and Answers." Computer Resource Security Center. March 5, 2001. <http://csrc.nist.gov/encryption/aes/round2/aesfact.html> (June 16, 2004).

Architectural Math

Overview

Architectural mathematics uses mathematical formulae and algorithms for designing various architectural structures. Most of these structures are buildings such as museums, galleries, sport complexes and stadiums, theatres, churches, cathedrals, offices, houses, and so on. Architectural mathematics is also used to design open spaces in cities and towns, recreational places including gardens, parks, playgrounds, water bodies such as lakes, ponds and fountains, and a variety of physical construction and development.

Put simply, architectural math pertains to mathematical concepts that are central to architecture. These architectural math concepts are also used in several other activities that we see in our daily lives. They are extensively used in sports, technology, design, aviation, medicine, astronomy, and much more.

Understanding architectural math requires knowledge of various two-dimensional as well as three-dimensional shapes such as square, rectangle, triangle, cube, cuboids, sphere, cone, and cylinder. There are also other concepts such as symmetry and proportion that are integral to architecture math. The pyramids of Egypt, for example, are based on these principles.

Fundamental Mathematical Concepts and Terms

As architectural designs are strongly inspired and implemented using various shapes and forms, basic architectural mathematics involves understanding underlying principles that derive such shapes and forms. This requires understanding geometric equations associated with its visual representations. In other words, architectural mathematics is not expressed as simple numbers but rather as graphical or visual forms. What follows are some of the most widely used architectural math concepts.

RATIO

A ratio is a comparison by division of two quantities expressed in the same unit of measure. In other words, you get a ratio by dividing two numbers of quantities. The ratio may be represented in words or in symbols. For example, if segment Line A is one inch long and Line B is two inches long, we say that the ratio of Line A to B is one to two. In terms of mathematical symbols, the ratio may be denoted in fractional form as ½, or it may be expressed as 1:2.

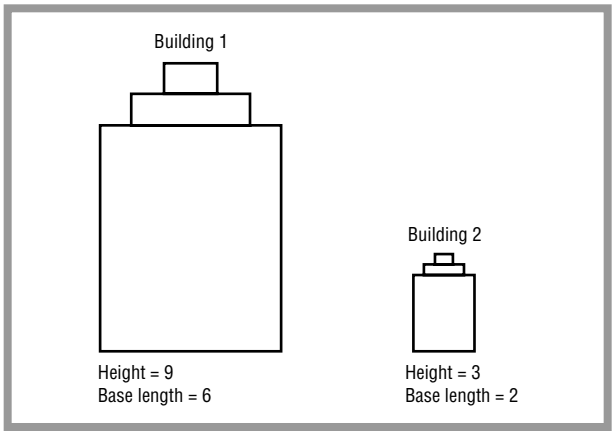


Figure 1.

Since ancient times, Greeks and Romans were known to be the most elaborate builders. They had a flair for architecture and built structures that were pleasing to the eye. They were convinced that architectural beauty was attained by the interrelation of universally valid ratios. Frequently, complicated mathematical ratios were used by architects to accomplish their goals. Take, for example, the golden ratio (or phi)—1.618. This ratio that has applications in many areas, has been extensively used in architecture—both modern and ancient.

PROPORTION

Proportion, like ratio has always been a vital component of architectural math. The ancient Greeks and Romans followed certain mathematical proportions (and ratio) to attain order, unity, and beauty in their buildings. Using simple mathematical formulae (based on proportion) they were able to establish a unique relationship among various parts of buildings. Such relationships have been used for generations.

To better understand the concept of proportion, consider an example of two buildings variable in height and base, however displaying the same proportion (see Figure 1).

In Figure 1, there are four terms that would define proportion of one building to another. These are 2, 3, 6, and 9. A proportion is an equation that states the ratios of comparison are equal. Thus, in the above example we would say that Building 1 is in proportion (or proportionate) to Building 2 if $6/9 = 2/3$, or $9/6 = 3/2$. This is the case, and hence the statement that the buildings are in proportion holds.

If the ratios for two objects are not equal, they would not be in proportion. Proportion can also be used to calculate the ratio of the total magnitude (in this case size) of the two objects. For example, in our case 9/6 can also be expressed as $3/2 \times 3$. Thus, we can say that Building 1 is three times the size of Building 2.

SYMMETRY

In architecture, one way to attain balance and response while designing structures is by the use of symmetry. Architecture is based on principles of balance. Basically, if most architectural forms are divided into two equal parts by a line in the center, the opposites sides of the dividing line would be similar (or even identical). This concept that can also be seen in all basic geometric forms (square, rectangle, circle, triangle, and so on) is known as symmetry and is extensively used architectural structures—modern and ancient.

See Figure 2 to understand symmetry further. In this figure (2a), triangle ABC and triangle BCD are symmetric about line m. The corresponding sides and corresponding angles of the triangles are similar. In other words, triangle CBD is the reflection of triangle ABC, and m is the line of symmetry. Such type of symmetry is also known as symmetry by reflection.

Forms can also be created using translation or sliding symmetry. This type of symmetry involves two or more similar forms of the same size and facing the same direction. In figure (2b), all images are similar to each other and face the same direction. Another way to understand this is imagine that the same image has been slid on the line p (and thus the name sliding symmetry).

The third method of attaining symmetry is by rotation. If a figure, after rotating it around a central point by less than 360°, remains unchanged, then it has rotation symmetry. For example, in figure 2c, if this form is rotated from the central point B by 180°, the resulting form would be the same. Thus, the figure has rotation symmetry for a rotation of 180°.

It is interesting to note that all geometric forms (square, rectangle, triangle, pyramid, hexagon, etc.) can have reflection, translation, and rotation symmetry. It is for this reason, they are used extensively in architecture.

SCALE DRAWING

For designing any structure (a building, a house, or a city), an architect is required to convert his ideas to drawings. These drawings provide homeowners, contractors, carpenters, and others with a small diagram of the final structure. The drawings show in detail the sizes, shapes,

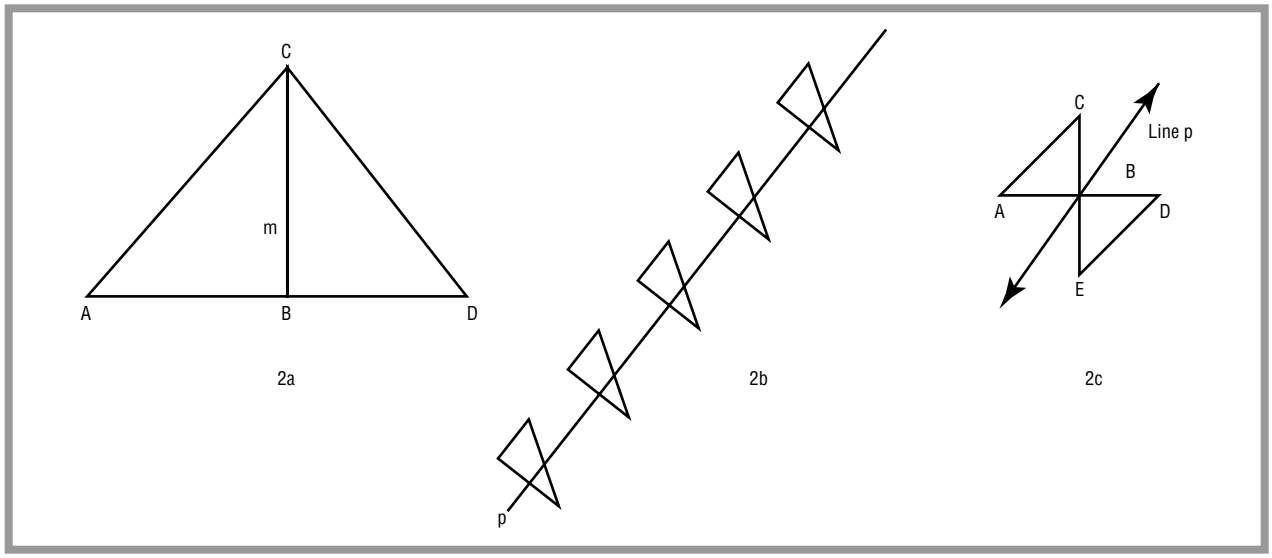


Figure 2.

arrangements of rooms, structural elements, windows, doors, closets, and other important details of construction. For example, a drawing for a house would specify the area (length, width, and height) of every room including the living area, bedroom, and bathroom at every floor. Such miniature reproductions of the structure are called scale drawings.

Scale drawings that represent parts of a structure must be in proportion to the actual structure.

To do this, architects use a specific scale corresponding to the actual size. For example, a scale such as ¼ inch = 1 foot, would suggest that a length of ¼ inch on the scale drawing is equal to one foot within the actual structure.

Scale drawings allow an architect to visualize a structure before building it.

MEASUREMENT

Measurement is important and is required while designing any building, right from the planning stage to the actual construction work. The instrument used to measure objects is the ruler, or a measuring tape. Architects, carpenters, and designers use measure-ments to come up with accurate scale drawings before starting construction work. Throughout the process of building any structure, measurement is extensively used.

Measurement can be expressed as inches, feet, and yards (English system), or centimeters, and meters (metric system).

A Brief History of Discovery and Development

There is a commonality between the seventeenth-century Round Tower of Copenhagen, the thirteenth-century Leaning Tower of Pisa, Houston's Astrodome, (the first indoor baseball stadium built in the United States), the vast dome of the Pantheon in Rome, a Chinese pagoda, and the Sydney Opera House. All these buildings were built using architectural math concepts such as scale, measurement, ratio, proportion, and symmetry.

Architectural mathematics has always been a vital part of structural design. The pyramids of Egypt used basic principles of the geometric "pyramid"—a square base and an apex tapering as the elevation of the pyramid increased. The pyramid shape provides higher stability compared to other structures as it is able to counter wind forces and natural forces, such as earthquakes, much more effectively than compared to most other shapes.

The same concept of visual geometry was used while constructing the Eiffel Tower in Paris. It has a square base with a narrowing apex as one moves toward the top of the structure. One key aspect of mathematics to be considered while building a pyramid structure is the use of ratio and proportion in designing the base and apex of the tower. Higher the ratio of base to the apex, the higher will be the stability of the structure to withstand the various forces. This has been kept in mind while designing the

Carpenters use a variety of everyday math skills. CORBIS.

above mentioned structures (and many other around the world).

The influence of mathematics on architecture and its principles can be seen since the time of the Greek mathematician Pythagoras (569 B.C.–475 B.C.). Pythagoras, and his followers, believed that all things could be represented in numbers. This concept has been used extensively in architecture, ever since.

Pythagoras, after conducting many experiments, found out that music depended considerably on mathematics. He concluded that musical scales (notes) depended on ratios of small integers. Architects adopted this principle and designed buildings based on ratios of smaller integers or units. These smaller units could be units of length, size, or dimension. For example, a simple wall consisted of smaller blocks equal in length.

Pythagoras also believed that all numbers could be represented as geometrical shapes. Furthermore, he developed an idea that geometrical symmetry based on proportion is far more appealing, visually. This is the very concept that architects used while designing buildings and other structures. The structures that were then built in ancient Greece and Rome were based on symmetry.

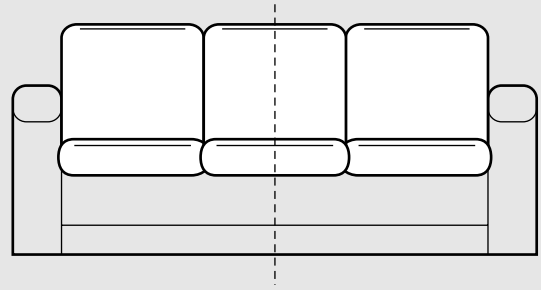
Architectural math concepts were eventually used in a variety of other areas including astronomy, carpentry, jewelry design, and more. It is not known how many of these concepts were used in architecture first and then later re-used in other fields of work. However, they are interlinked and there is a high possibility that these concepts inspired other designers and engineers to use mathematics effectively to justify the form or function of other objects and devices, just the way architects used it to express their building designs.

Real-life Applications

ARCHITECTURE

It is evident by now that architecture is the most common application of architectural mathematics. These concepts are used in a number of ways by architects the world over.

Carpentry



"Measure twice and cut once" is a principle slogan of carpentry. Cutting out pieces of wood in such a way that they subsequently fit together to make a beautiful cabinet, desk, cupboard or whatever calls for correct measurements of length, width and height. The manufacture of these and other items calls for the correct measurement of slope; no one wants to try to eat off of a table that is so slanted that the soup spills onto the floor! Nailing pieces of wood together calls for a mathematical distinction between perpendicular and an angle less than 90 degrees.

A carpenter needs to understand the size and proportion of each object depending on the person who is expected to use it. These are again based on ergonomic standards (see the section on Ergonomics). Subsequently, the designs reflect most of the basic concepts of architectural mathematics.

For example, table tops often have wood pieces cut and put together to form a symmetric design, exemplifying their beauty. Symmetries and ratios are clearly evident in the design of a couch as well, which is based on bilateral symmetry (see figure at above right).

The myriad number of carpentry processes that go into the construction of a house are rooted in math. Carpenters get involved in house construction following the laying of the foundation. Construction of a floor frame is essential. Often the wood frame sits directly on the foundation, and is not fixed or bolted to the foundation. The weight of the house will provide the force to keep the frame intact. But, for this operation to be successful, the frame must be the same length, width and shape as the foundation and have enough cross braces to provide strength. Proper measuring is crucial, as is the fitting together of the slabs of wood that make up the skeleton of the frame.

The cross braces, which are also called "joists" are attached to a center beam that runs down the center of the house. Once again, proper positioning of the beam is essential to establishing the support needed for the floors to come.

Once a floor is installed on the floor frame, walls can be built. This construction is fraught with measurements. For example, since special vertical supports need to be in the right places to accommodate the interior walls. Other side supports are usually positioned 8 to 16 inches apart and comprise the supporting studs. Doors, windows and other exterior openings must be properly located. In the case of windows, a special structure called a header needs to be built above the window opening. It will give the wall enough strength over the expanse of glass to support the roof. Typically, this phase of the construction requires detailed plans of the structure, with accurate measurements. A blueprint of the project is a necessary and prudent tool.

USE OF SPACE

Architecture is all about using physical three-dimensional space. The ways in which this space is used is one of the most important aspects of architecture. Many mathematical concepts are visible in architectural designs that include spaces. For example, a building design may leave an open space in the center, which is often geometrical in shape such as a square, rectangle, or a circle, with several other geometric designs surrounding the space. Spaces are always designed based on mathematical principles or ratio, proportion, and symmetry.

Designs based on symmetry and space can be seen in several open spaces and squares. One great example of the same is St. Peter's Square in Vatican City, Rome, where a circular symmetry of elements is focused on the central piece. The intersecting lines in the circle have a focal point of concentration, a place where people can gather around.

USE OF GRIDS

Elements in a building are often arranged forming a grid that resembles a set of parallel and perpendicular lines on a piece of paper. Again these elements are based on symmetry and patterns.

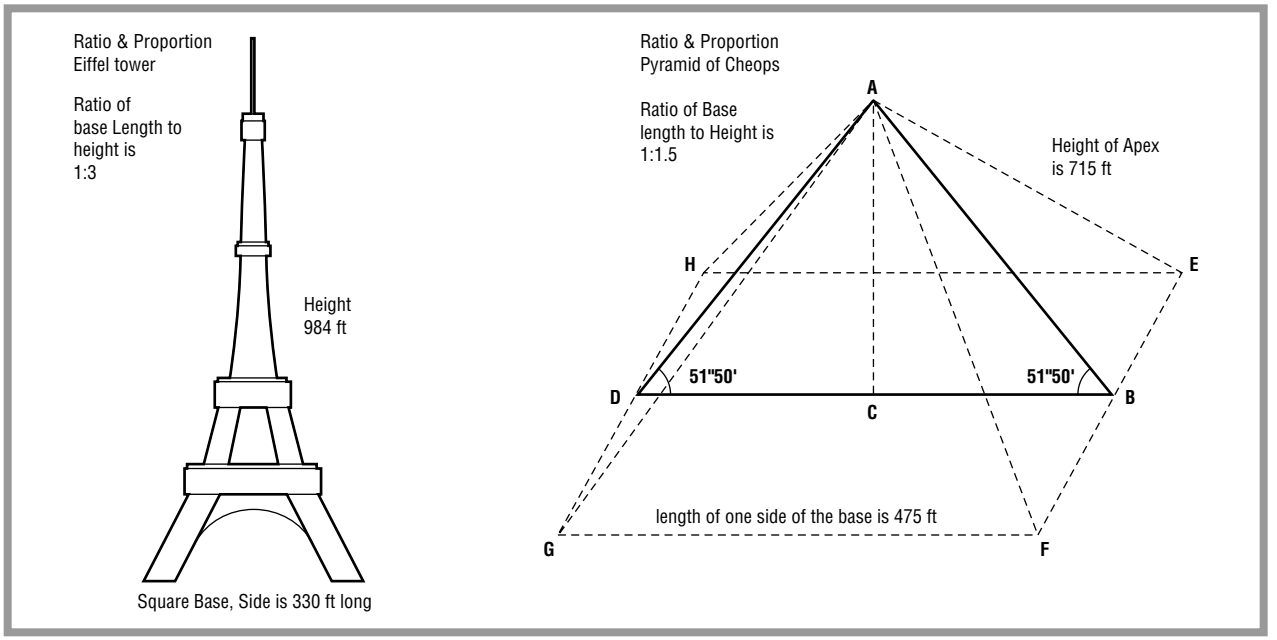


Figure 3.

USE OF RATIO AND PROPORTION

Historical monuments and modern buildings, alike, have used architectural mathematics extensively. This is reflected in several landmarks. As stated earlier, the Pyramids of Egypt are a very good example of the use of a simple form (the geometric pyramid shape), having a square base and tapering as its height increases. These are built such that their base and height have specific ratio and proportion. This is done to impart greater strength and stability. The same is the case with the Eiffel tower in Paris.

For example, the ratio of the base of the Pyramids with their height is almost 1:1.5. The ratio gives the structure higher stability. Conversely, the Eiffel tower has a ratio of 1:3 for base to height. Both these structures have different ratios. However, they both impart stability to the structure due to their shape and design. (See Figure 3.)

Architectural designs also use ratio and proportion to justify the dimensions of elements within the buildings. This includes length and width of the corridor, its proportion to doors that lie within the corridor, the height of the ceiling from the floor with respect to the type of building, and the ratio of size of the steps on a staircase with respect to the total height of the staircase—all these aspects are considered vital while designing a structure.

USE OF ARCHITECTURAL SYMMETRY IN BUILDINGS

Reflection symmetry (also known as bilateral symmetry) is the most common type of symmetry in architectural designs. In bilateral symmetry, the halves of a composition mirror each other. Such symmetry exists in the Pantheon in Rome. We find the same symmetry in the mission-style architecture of the Alamo in San Antonio, Texas.

Bilateral symmetry existed in several buildings built during the Roman or Greek periods. Modern architects also use such symmetry widely for various structures (see Figure 4).

Additionally, translation and rotation symmetry are also employed considerably in modern architectural designs.

USE OF RECTANGLE AS "GOLDEN RECTANGLE" AND "GOLDEN RATIO"

Since ancient times, architectural designs have used the golden ratio (1.618) in various ways. One of the best examples being the Parthenon, the main temple of the goddess Athena, built on the Acropolis in Athens. The front of the Parthenon is a triangular area that fits inside a rectangle whose sides are equivalent to the golden ratio (the rectangle is popularly known as the golden rectangle). The golden ratio and its related figures were incorporated into every piece and detail of the Parthenon.

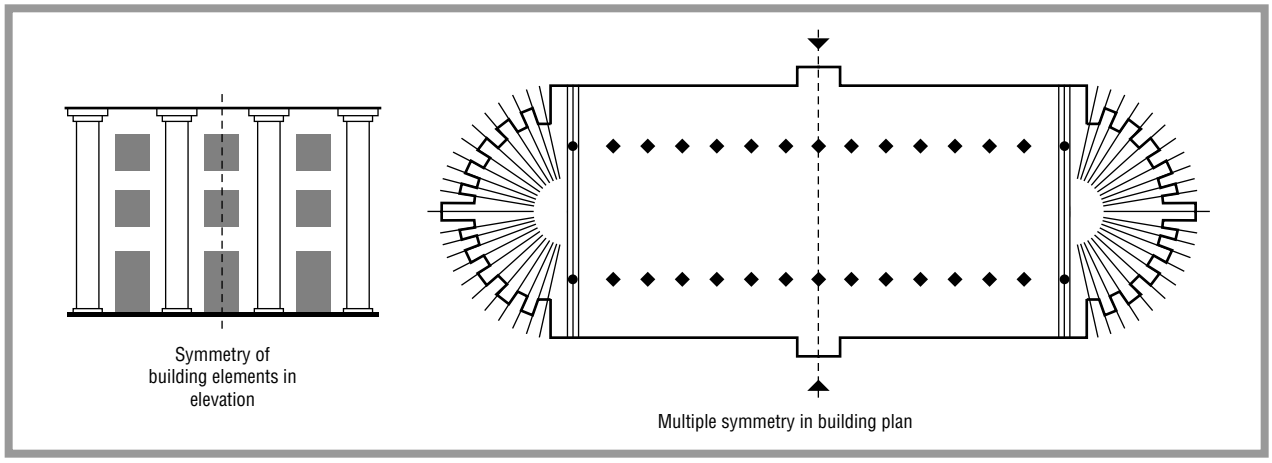


Figure 4.

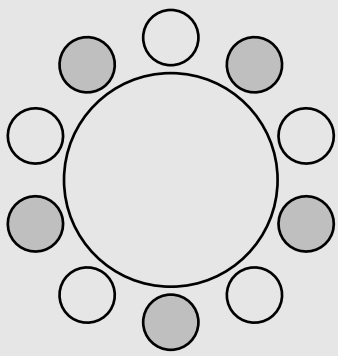
The same math principles that allowed the construction of the Arch of Constantine in Rome also allow designers to shape modern home interiors. The arch distributes load. TRAVELSITE/DAGLI ORTI. REPRODUCED BY PERMISSION.

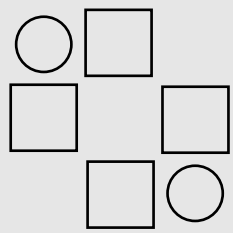
Decorating

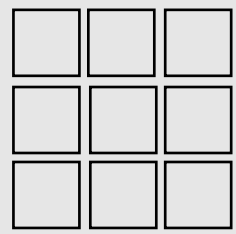
Numerous symmetrical shapes and forms are used while decorating furniture and home accessories ranging from a flower vase to the kitchen sink. The use of architectural shapes and concepts is clearly visible in every decorative aspect of the complete design.

Besides the arrangement of these shapes and forms (see figures below), the concept of symmetry also plays a vital role in the layout of these. For example, while decorating a room, most interior designers would ensure that the entire layout of the room (and how all elements within the room are placed) is based on symmetry principles. The main purpose is to give a decorative touch to the room to make it visually more appealing.

Home accessories, especially decorative artwork (vase, glassware, china pottery, and so on) are often made of wood or ceramics. These decorative pieces more often than not are also based on principles of symmetry. Their shape, exterior designs, and colors are amazingly symmetrical, and in many cases are based on basic geometric forms and shapes—much like designs in architecture.







The Triumphal Arch of Constantine, and the Colosseum—an amphitheater in Rome built in around A.D. 75 (both in Rome)—are other great examples of ancient use of golden relationships in architecture. The main idea behind employing this ratio was to make the structure visually appealing and also more stable.

USE OF BASIC FORMS AND SHAPES OF GEOMETRY

Apart from mathematical concepts such as ratio, proportion, and symmetry, most architectural designs are based on basic geometric shapes and forms including triangle, rectangles, pyramids, cones, cylinders, and more. Although, when viewed as a whole these structures would have basic shapes, their interiors can always be represented by the above mentioned mathematical concepts.

The Taj Mahal in India is an example of the use of a basic shape or form—the cube. The Taj Mahal was built as a cube, where the four minarets and the center burial tomb of the queen all are contained in a perfect cube. The length, breadth, and height of all sides are equal in dimension. Additionally, the sense of ratio, proportion, and symmetry of this structure is precise and spell-binding.

A modern example of the use of basic shapes is the Pentagon, in Washington, D.C. The Pentagon's five sides are equal in length, denoting a perfect pentagon. Within the main structure, there are five concentric pentagons of corridors and offices. Again, these internal pentagons are symmetric and in proportion to each other.

SPORTS

Geometric shapes and forms, symmetry, ratio, and proportion have found a place in sports as well. Practically, every field sport uses architectural math principles. A tennis court, basketball court, football, hockey, soccer

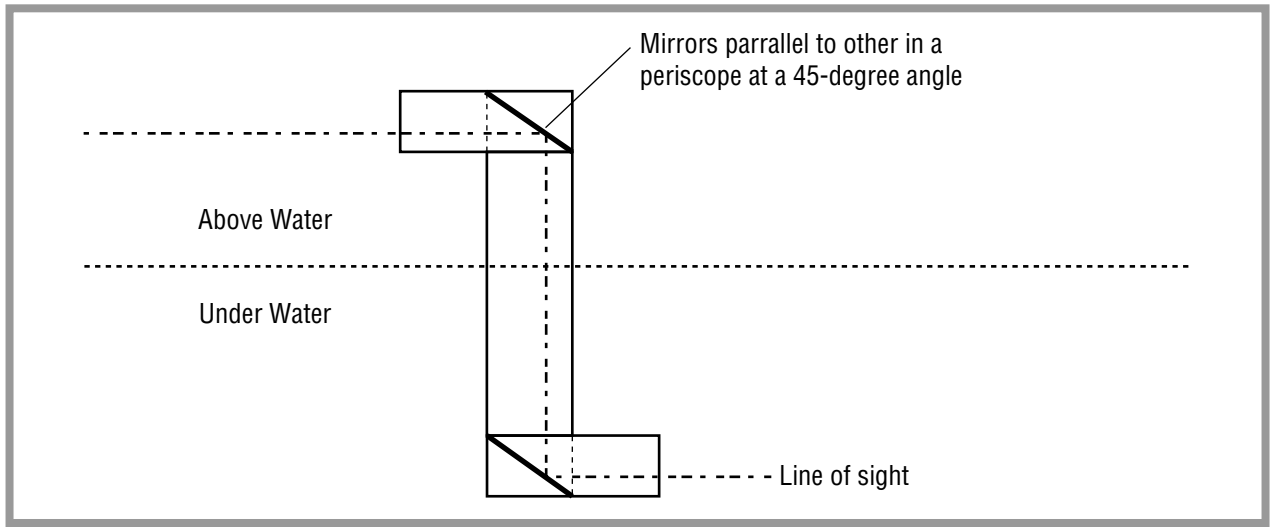


Figure 5.

fields—all of them are rectangles having a center line dividing each into two halves for each team or players. These are perfect examples of reflection symmetry. Besides, many of these sports have fields that have sides forming the "Golden Rectangle." The ratio of a side to its length is based on the golden ratio—a concept adopted by architects to depict buildings during the Roman and Greek periods, as discussed earlier.

In addition, other concepts of mathematics that are commonly applicable to architectural designs, such as measurements and scales also apply to field sports.

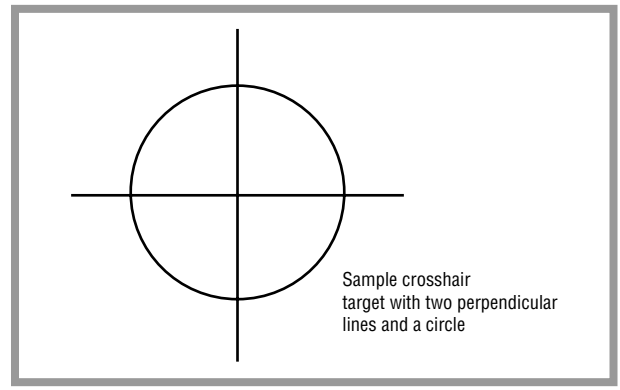


Figure 6.

TECHNOLOGY

Technology tools and devices use architectural math concepts of symmetry and proportion to facilitate their underlying functions. Equipments such as a periscope used in submarines, guns in aircrafts, and satellite transmission use the principle of symmetry.

A periscope is commonly used in submarines. It is a device that can help view objects such as ships and other water vessels above the water surface, while still being underwater. A periscope has two mirrors placed at a 45° angle to the eye's line of sight along with another mirror placed at 45° parallel to the first one at a variable height (see Figure 5). This allows a person to view objects using the laws of reflection at different heights while maintaining his/her own position and eye level.

The principle of periscope is clearly visible in the rotational symmetry concept of many architectural building designs. The use of two mirrors can be compared with two parallel lines to reflect a design and make it look symmetrical. The concept is the same, however its application and use varies drastically.

Fighter aircraft guns are often assisted by the visual cross-hair—two perpendicular lines, where the point of intersection is often pinpointing or locating a target (see Figure 6). There are several aircraft guns that have two cross-hairs, one for the aircraft gun itself, and the other for the object. Once the two cross hairs coincide with each other, or are symmetrically aligned with each other, the target object is in line with the gun point. In other words, the crosshairs are now pointing at the target. The target can be locked and shot. This entire mechanism is based on principles of symmetry.

Another such technological tool is the fan. A fan works on the principles of architectural math, mainly symmetry, and is used in several applications including aircrafts, helicopters, wind mills, and air conditioners, as well as industrial/home establishments such as kitchens

(exhaust fans). A fan has typically three wings, which are similar in size, shape, ratio, and proportion. On rotation, they (wings) generate air and help run several mechanical as well as electrical devices. A simple fan is one of the most commonly used mathematical applications of symmetry and geometric shape (circle).

USING SYMMETRY IN CITY PLANNING

Since the early 1930s, most cities in the world have developed in similar ways. City planners have always focused on symmetrical models for planning a new city or even developing existing cities further. Many cities have a central area known as the central business district (CBD). This is where businesses within a city are concentrated. The areas around the CBD are mostly residential.

The manner in which residential areas have developed over the years is comparable (around the world). A city can be thought of as a group of clusters, where each cluster comprises of a number of buildings, roads, and other structures. The entire city consists of numerous such clusters arranged symmetrically. In other words, a city would consist of a central area (CBD) and several similar clusters around the CBD placed in a symmetrical pattern. This concept is based on the principle of translation symmetry (also known as a fractal or motif).

Some of the biggest cities in the world, including New York, London, Paris, Beijing, and so on are in many ways based on the architectural mathematical concept of symmetry. That said, the late 1990s and early 2000s have started witnessing newer cities that are far more decentralized. In other words, the concept of a central business area and symmetrical clusters of residential areas around it is losing popularity. Clusters within some cities have become dispersed and random rather than symmetric.

ERGONOMICS

Ergonomics is a science that studies technology and how well it suits the human body. Ergonomics involves understanding basic body parts, their functions and abilities to operate equipments, machinery, products, and other technological devices. Ergonomics is commonly used while designing cars, among other things. Ergonomic car designs are based on the principles of ratio and proportion. In other words, car designers use principles and math concepts that are used considerably in architecture to come up with designs that better fitand serve the human body.

For example, the height from the surface, inclination, and movements patterns in a car seat for drivers are all designed in proportion to the human body. The ratios are extremely critical here. The size of the seat has to be in proportion with the size of an average human driver. Besides, you do not expect a person to have a giant wheel in front of him/her, the size of the wheel (the diameter of the wheel) has to be in proportion to the size of the hand grip, shoulder width, and distance between the wheel and person driving the car. All these elements are carefully incorporated into the design of cars.

Similarly, interior designers also use ratios to design various objects (such as beds, tables, chairs, and so on) within a house. These ratios are based on ergonomic standards. For example, a bed is designed such that it is in proportion to the human body. In Sweden, beds have a length of 7 feet (2.1 m), while beds in Japan are rarely 6 feet (1.8 m) long. This is due height differences in the populations. The average height of an individual in Japan is 5 feet 2 in (1.5 m), while the average height of a person in Sweden is about six feet (1.8 m). This also influences other design standards such as height of the bed from the floor, width of the bed, and portability of the bed.

The size of the window is also often based on the proportion of human body. A window in a house will be smaller, compared to a window in a public building. The proportion of both windows may be same implying that the ratio of their width/height is equal. However, their sizes would differ.

Architectural mathematical concepts such as ratio and proportion form an integral part of ergonomics, especially when it comes to design related issues.

JEWELRY

Ornaments made of gemstones, diamonds, gold, and silver use symmetry of arrangement extensively. A cut of a diamond often displays several shapes and forms. Gold is molded into several geometric forms to add value to an ornament.

Consider, for example, pendants that are more often than not designed using principles of symmetry. The symmetry in such pendants is visible in architectural structures as it is in nature (arrangement of flowers and fruits on trees). Ornament designs are often very intricate and require a finer view to understand their symmetry, ratio, and geometric shapes. Such symmetric designs are not limited to pendants but are also visible in rings, bracelets, and several other ornaments.

Compare these with symmetric designs in the ceilings of several domes and museum galleries and a stark resemblance is clearly evident. Mirrors, stained glass, and other shinny materials that are commonly used to signify architectural designs in building interiors are very similar to ornament designs—with respect to their visual arrangement and their underlying mathematical principles.

Cathedrals are a common example, where the architecture is inspired by arranging materials and objects in symmetry, similar to that in ornaments and jewelry.

ASTRONOMY

Fundamentals of architectural math including distance, size, and proportion are also visible in various astronomical advancements. The telescope is one such example. Telescopes are used to view stars and planets located in far away galaxies. The distance is measured in light years. The distance traveled by light in one year is known as a light year (light travels at a speed of 186,000 miles per second). This gives an indication of the distances between the Earth and some of the stars and planets.

Telescopes are used to magnify the image of these objects. This is done by using different lenses. Larger telescopes, such as the Hubble telescope, are able to magnify objects situated at a larger distance. Smaller telescopes in comparison have lower magnification implying lower visibility and clarity.

One of the basic mathematical principles of telescopes is scaling—a concept extremely common in architecture. Just like architects draw scale diagrams using ratio and proportion, telescopes use the same principles to magnify objects situated at large distances. In other words, telescopes present a scale model of an object that is not otherwise visible (or too tiny) with the naked eye.

Although, larger telescopes magnify objects that are further away, as compared to the smaller telescopes, the degree of magnification (of both types of telescopes) is always in proportion.

TEXTILE AND FABRICS

Cloth or fabrics are used for a variety of purposes. This includes bed sheets, covers, clothes, apparels, wipes, and more. Fabrics are textile products that require knitting. These are made from fibers of cotton, nylon, or other types. However, most of these fabrics do not have any value until a design is printed or woven on them. In other words, fabric prints carry considerable value to a plain piece of fabric or cloth. People would usually buy fabrics with visually appealing prints, rather than those that are plain.

Symmetry, which is used commonly used in architecture, is often reflected in fabric or cloth designs. Most fabric designs are composed of motifs. Motifs are repetitive use of a single design concept, style, or shape—Motifs signify symmetry (translation symmetry). The type of motifs could range from a leaf or a flower of same color or style repeated over the entire fabric print. The design varies depending on the final use of the fabric. Bed sheets, clothes, fashion apparels, and so on have different symmetry motifs depending the type of fabric, their manufacturing price, and quality of print.

Motifs are not limited to floral or color patterns but are often extend to lines, simple geometric shapes (squares, circles, rectangle, etc.), blocks, and much more. In some cases, once the fabric is cut or is stitched to make the final product, the symmetry may be lost. Nevertheless, the design is still based on the very principle of symmetry.

This is one of the most common applications in daily life that uses mathematical concepts of architecture in a very different way.

ARCHITECTURAL CONCEPTS IN WHEELS

Commuting has become an integral part of our daily life. We drive (on our own or in public transportation) to work, to school, to attend meetings, to go shopping or buy groceries. We require transportation to reach different places. Today, transportation is seen as a necessity.

Transportation is facilitated by public buses, railways, airplanes, and cars. All of these use wheels. A wheel, be it of rubber, magnet, or iron, is a vital component of any automobile. The wheel consists of a bar in its center known as the axle. The width of the axle is governed by the width of the carriage (weight of the automobile) required. Subsequently, the width varies in trains, buses, and cars. While designing wheels, engineers must ensure that the size of the wheel and the axle is in proportion to the total weight of the vehicle (including the people it carries) as well as the speed at which the vehicle can travel.

Ratio and proportion play a very important role in defining the diameter, width and the number of wheels that have to be attached with a vehicle. Higher the load to be carried, the more number of wheels (and even stronger wheels) will be required. Similarly, the longer the length of the vehicle, more the number of wheels required. Airplanes do not travel on wheels but require them to land and take off. However, the proportion of their wheels is much greater when compared with other vehicles as the amount of load is much higher. Besides, the size of the plane is also much larger when compared with other vehicles.

In short, wheels have to compliment the size of the vehicle and its intended purpose. Automobile design uses mathematical concepts of ratio and proportion, similar to those used in architecture. These are also based on ergonomical standards (see section on Ergonomics).

Key Terms

Proportion: An equality between two ratios.

Ratio: The ratio of a to b is a way to convey the idea of relative magnitude of two amounts. Thus if the number a is always twice the number b, we can say that the ratio of a to b is "2 to 1." This ratio is sometimes written 2:1. Today, however, it is more common to write a ratio as a fraction, in this case 2/1.

Scale: The ratio of the size of an object to the size of its representation.

Symmetry: An object that is left unchanged by an operation has a symmetry.

Where to Learn More

Books

Rossi, Corinna. *Architecture and Mathematics in Ancient Egypt.* Cambridge University Press, 2004.

Williams, Kim. *Nexus III: Architecture and Mathematics.* Pacini Editore, 2000.

Web sites

University College London, Department of Geography. "Fractals New Ways of Looking at Cities" <http://www.geog.ucl.ac.uk/casa/nature.html>(April 9, 2005).

Yale New Haven Teachers Institute. "Some Mathematical Principles of Architecture" <http://www.cis.yale.edu/ynhti/curriculum/units/1983/1/83.01.12.x.html> (April 9, 2005).

Area

Overview

An area is a measurement of a defined surface such as a face, plane, or side. Conceptually, an object's area can be compared quantitatively to the amount of paint needed to cover the object completely. However, in contrast to measures of volume in pints, liters, or gallons, area measurements are expressed in units such as square feet, square meters, or square miles. Calculations of area are basic to science, engineering, business, buying and selling land, medicine, and building.

Fundamental Mathematical Concepts and Terms

AREA OF A RECTANGLE

Every real-world object and every geometrical figure that is not a point or a line has a surface. The amount or size of that surface is the object's or figure's area. There are many standard formulas for calculating areas, the simplest and most commonly used being the formula for the area of a rectangle. To find the area of a rectangle, first measure the lengths of its sides. If the rectangle is W centimeters (cm) wide and H cm high, then its area, A, is given by $A = W\text{ cm} \times H\text{ cm}$.

Centimeters are used here only as an example. The units used to measure length—centimeters, inches, kilometers, miles, or anything else—do not change the basic formula: area equals width times height. So, for example, a typical sheet of typing paper, which is 8.5 inches wide and 11 inches high, has area $A = 8.5\text{ inches} \times 11\text{ inches} = 93.5$ square inches.

UNITS OF AREA

Area has now been explained in terms of "square inches" (or centimeters). This means that on the right-hand side of the formula $A = W\text{ cm} \times H\text{ cm}$, four terms are multiplied: W, H, and cm (twice). These four terms can be reordered to give $W \times H \times \text{cm} \times \text{cm}$. It is customary in mathematics to use the square notation when a term is multiplied by itself, so cm × cm is always written cm^2, which is centimeters squared, or square centimeters. Another way of writing the rectangle area formula is, therefore, $A = WH\ \text{cm}^2$. Area is therefore measured in units of square centimeters—or square inches, square feet, square kilometers, square miles, or any other length measure squared. For example, a square with edges 1 foot long has an area of 1 square foot.

When talking about physical materials such as cloth, land, sheet steel, plywood, or the like, it is important to

Areas of geometric shapes

Geometric figure	Dimensions	Formula for area
rectangle	width W, height H	$A=WH$
square	side length H	$A=H^2$
circle	radius R	$A=R^2$
triangle	base B, height H	$A=1/2\ BH$
parallelogram	base B, height H	$A=BH$
trapezoid	base B, top T, height H	$A=1/2\ (B+T)H$

Figure 1.

give correct units for length and area. However, in mathematics it is common to not use units. The norm is to say that an imagined rectangle has a length of 4, a height of 5, and an area of $4 \times 5 = 20$.

AREAS OF OTHER COMMON SHAPES

The simplest rectangle is a square, which is a rectangle whose four sides are all of equal length. If a square has sides of length H, then its area is $A = H \times H = H^2$.

The standard formulas for finding the areas of other simple geometric figures are depicted in Figure 1.

Notice that in all the area formulas, two measures of length are multiplied, not added. This means that whenever an object is made larger, its area increases faster than its height or width. For example, a square that has sides of length 2 has area $A = 2^2 = 4$, but a square that is twice as tall, with sides of length 4, has area $A = 4^2 = 16$, which is four times larger. Likewise, making the square three times taller, with sides of length 6, makes it area $A = 6^2 = 36$, which is nine times larger. In general, a square's area equals its height squared; therefore its area "increases in proportion to" or "goes as" the square of the side length. Consequently, a common rule of thumb for sizes and areas is, increasing the size of a flat object or figure makes its area grow in proportion to the square of the size increase.

AREAS OF SOLID OBJECTS

Three-dimensional objects such as boxes or balls also have areas. The area of a box can be calculated by adding up the areas of the rectangles that make up its sides. For example, the formula for the area of a cube (which has squares for sides) is just the area of one of its sides, H^2 multiplied by the number of sides, which is 6: $A = 6H^2$.

Calculating the area of a rounded object like a ball is not as simple, because it has no flat sides and none of the standard formulas for simple geometric shapes can be used to find the areas of parts of its surface. Fortunately, standard formulas were worked out centuries ago for simple rounded objects like cones, spheres, and cylinders; these formulas are listed in many math books. For example, the area of a sphere of radius R is $A = 4\pi R^2$ (π, pronounced "pie," is a special number approximately equal to 3.1416; see the article on "Pi" in this book). The Earth, which is basically sphere-shaped, has an average radius of 6,371 kilometers (km), or about 3,956 miles. Its surface area is therefore $A = 4\pi 6{,}371^2 = 510{,}060{,}000\ \text{km}^2$, which is about 316,750,000 square miles. The Earth is 53 times the area of the United States.

A Brief History of Discovery and Development

The calculation of areas was one of the earliest mathematical ideas to be developed by ancient civilizations, preceded only by counting and length measurement. The ability to calculate areas was originally needed in the buying and selling of land. Four thousand years ago the Egyptian and Babylonian civilizations also knew how to calculate the area of a circle, having worked out approximate values for the number π. The ability to calculate areas was also useful in construction projects. The pyramids of Egypt, for example, could only have been constructed with the help of sophisticated geometric knowledge, including formulas for the areas of basic shapes. Calculation of the areas of spheres and other solid objects also dates back to the ancient Egyptian and Babylonian civilizations. Similar knowledge was discovered independently by Chinese mathematicians at about the same time.

In the seventeenth century, the calculation of the areas of shapes with smoothly curving boundaries was an important goal of the inventors of the branch of mathematics known as calculus, especially the English physicist Isaac Newton (1642–1727) and the German mathematician Gottfried Wilhelm von Leibniz (1646–1716). One of the two basic operations of calculus, integration, describes the area under a curve. (To understand what is meant by the area under a curve, one must imagine looking at the flat end of a building with an arch-shaped roof. The area of the wall at the end of the building is the area under the curve marked by the roofline.) The area under a curve may stand for a real physical area—if, for example, the curve describes the edge of a piece of metal or a plot of land—or, it may stand for some other quantity, such as money earned, hours lived, fluid pumped, fuel consumed, energy generated. The extension of the area concept through calculus over the last three centuries has made modern technology possible.

About 70% of the surface area of Earth is covered with water. U.S. NATIONAL AERONAUTICS AND SPACE ADMINISTRATION (NASA).

Real-life Applications

DRUG DOSING

The amount of a drug that a person should take depends, in general, on their physical size. This is because the effect of a drug in the body is determined by how concentrated the drug is in the blood, not by the total amount of drug in the body. Children and small adults are therefore given smaller doses of drugs than are large adults. The size of a patient is most often determined by how much the patient weighs. However, in giving drugs for human immunodeficiency virus (HIV, the virus that

causes AIDS), hepatitis B, cancer, and some other diseases, doctors do not use the patient's weight but instead use the patient's body surface area (BSA). They do so because BSA is a better guide to how quickly the kidneys will clear the drug out of the body.

Doctors can measure skin area of patients directly using molds, but this is practical only for special research studies. Rather than measuring a patient's skin area, doctors use formulas that give an approximate value for BSA based on the patient's weight and height. These are similar in principle to the standard geometric formulas that give the area of a sphere or cone based on its dimensions, but less exact (because people are all shaped differently). Several formulas are in use. In the West, an equation called the DuBois formula is most often used; in Japan, the Fujimoto formula is standard. The DuBois formula estimates BSA in units of square meters based on the patient's weight in kilograms, *Wt*, and height in centimeters, *Ht*: BSA = $.007184 Wt^{.425} Ht^{.725}$

In recent years, doctors have debated whether setting drug doses according to BSA really is the best method. Some research shows that BSA is useful for calculating doses of drugs such as lamivudine, given to treat the hepatitis B virus, which is transmitted by blood, dirty needles, and unprotected sex. (Teenagers are a high-risk group for this virus.) Other research shows that drug dosing based on BSA does not work as well in some kinds of cancer therapy.

BUYING BY AREA

Besides addition and subtraction to keep track of money, perhaps no other mathematical operation is performed so often by so many ordinary people as the calculation of areas. This is because the price of so many common materials depends on area: carpeting, floor tile, construction materials such as sheetrock, plywood, exterior siding, wallpaper, and paint, whole cloth, land, and much more. In deciding how much paint it takes to paint a room, for example, a painter measures the dimensions of the walls, windows, floor, and doors. The walls (and ceiling or floor, if either of those is to be painted) are basically rectangles, so the area of each is calculated by multiplying its height by its width. Window and door areas are calculated the same way. The amount of area that is to be painted is, then, the sum of the wall areas (plus ceiling or floor) minus the areas of the windows and doors. For each kind of paint or stain, manufacturers specify how much area each gallon will cover, the spread rate. This usually ranges from 200 to 600 square feet per gallon, depending on the product and on the smoothness of the surface being painted. (Rough surfaces have greater actual surface area, just as the lid of an egg carton has more surface area than a flat piece of cardboard of the same width and length.) Dividing the area to be painted by the spread rate gives the number of gallons of paint needed.

FILTERING

Surface area is important in chemistry and filtering because chemical reactions take place only when substances can make contact with each other, and this only happens on the surfaces of objects: the outside of a marble can be touched, but not the center of it (unless the marble is cut in half, in which case the center is now exposed on a new surface). Therefore a basic way to take a lump of material, like a crystal of sugar, and make it react more quickly with other chemicals is to break it into smaller pieces. The amount of material stays the same, but the surface area increases.

But don't larger cubes or spheres have more surface area than small ones? Of course they do, but a group of small objects has much more surface area than a single large object of the same total volume. Imagine a cube having sides of length L. Its area is $L = 6L^2$. If the cube is cut in half by a knife, there are now two rectangular bricks. All the outside surfaces of the original cube are still there, but now there are two additional surfaces—the ones that have appeared where the knife blade cut. Each of these surfaces is the same size as any of the cube's original faces, so by cutting the cube in half there has added $2L^2$ to the total area of the material. Further cuts will increase the total surface area even more.

Increasing reaction area by breaking solid material down into smaller pieces, or by filling it full of holes like a sponge, is used throughout industrial chemistry to make reactions happen faster. It is also used in filtering, especially with activated charcoal. Charcoal is solid carbon; activated charcoal is solid carbon that has been treated to fill it with billions of tiny holes, making it spongelike. When water is passed through activated charcoal, chemicals in the water stick to the carbon. A single teaspoonful of activated charcoal can contain about 10,000 square feet of surface area (930 square meters, the size of an American football field). About a fourth of the expensive bottled water sold in stores is actually city tap water that has been passed through activated charcoal filters.

CLOUD AND ICE AREA AND GLOBAL WARMING

Climate change is a good example of the importance of area measurements in earth science. For almost 200 years, human beings, especially those in Europe, the

United States, and other industrialized countries, have been burning massive quantities of fossil fuels such as coal, natural gas, and oil (from which gasoline is made). The carbon in these fuels combines with oxygen in the air to form carbon dioxide, which is a greenhouse gas. A greenhouse gas allows energy from the Sun get to the surface of the Earth, but keeps heat from escaping (like the glass panels of a greenhouse). This can melt glaciers and ice caps, thus raising sea levels and flooding low-lying lands, and can change weather patterns, possibly making fertile areas dry and causing violent weather disasters to happen more often. Scientists are constantly trying to make better predictions of how the world's climate will change as a result of the greenhouse effect.

Among other data that scientists collect to study global warming, they measure areas. In particular, they measure the areas of clouds and ice-covered areas. Clouds are important because they can either speed or slow global climate change: high, wispy clouds act as greenhouse filters, warming Earth, while low, puffy clouds act to reflect sunlight back into space, cooling Earth. If global warming produces more low clouds, it may slow climate change; if it produces more high wispy clouds, it may speed climate change. Cloud areas are measured by having computers count bright areas in satellite photographs.

Cloud areas help predict how fast the world will get warmer; tracking ice area helps to verify how fast the world has already been getting warmer. Most glaciers around the world have been melting much faster over the last century—but scientists need to know exactly how much faster. To find out, they first take a satellite photo of a glacier. Then they measure its outline, from which they can calculate its area. If the area is shrinking, then the glacier is melting; this is itself an important piece of knowledge. Scientists also measure the area of the glacier's accumulation zone, which is the high-altitude part of the glacier where snow is adding to its mass. Knowing the total area of the glacier and the area of the accumulation zone, scientists can calculate the accumulation area ratio, which is the area of the glacier's accumulation zone divided by its total area. The mass balance of a glacier—whether it is growing or shrinking—can be estimated using the accumulation area ratio and other information.

CAR RADIATORS

Chemical reactions are not the only things that happen at surfaces; heat is also gained or lost at an object's surface. To cool an object faster, therefore, surface area needs to be increased. This is why elephants have big ears: they have a large volume for their body surface area, and their large, flat ears help them radiate extra heat. It is also why we hug ourselves with our arms and curl up when we are cold: we are trying to decrease our surface area. And it is how cars engines are kept cool. A car engine is supposed to turn the energy in fuel into mechanical motion, but about half of it is actually turned into heat. Some of this heat can be useful, as in cold weather, but most of it must simply be expelled. This is done by passing a liquid (consisting mostly of water) through channels in the engine and then pumping the hot liquid from the engine through a radiator. A radiator is full of holes, which increase its surface area. The more surface area a radiator has, the more cool air it can touch and the more quickly the metal (heated by the flowing liquid inside) can get rid of heat. When the liquid has given up heat to the outside world through the large surface area of the radiator, the liquid is cooler and is pumped back through the engine to pick up more waste heat. Car designers must size radiator surface area to engine heat output in order to produce cars that do not overheat.

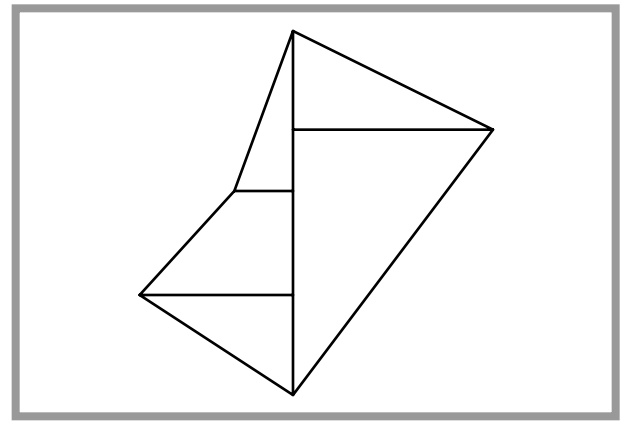


Figure 2.

SURVEYING

If a parcel of land is rectangular, calculating its area is simple: length × width. But, how do surveyors find the area of an irregularly shaped piece of land—one that has crooked boundaries, or maybe even a winding river along one side?

If the piece of land is very large or its boundaries very curvy, the surveyor can plot it out on a map marked with grid squares and count how many squares fit in the parcel. If an exact area measurement is needed and the parcel's boundary is made up of straight line segments, which is usually the case, the surveyor can divide a drawing of the piece of land into rectangles, trapezoids, triangles. The area of each of these can be calculated separately using a standard formula, and the total area found as the

sum of the parts. Figure 2 depicts an irregular piece of property that has been divided into four triangles and one trapezoid.

Today, it is also possible to take global positioning system readings of locations around the boundary of a piece of property and have a computer estimate the inside area automatically. This is still not as accurate as an area estimate based on a true survey, because global positioning systems are as yet only accurate to within a meter or so at best. Error in measuring the boundary leads to error in calculating the area.

SOLAR PANELS

Solar panels are flat electronic devices that turn part of the energy of sunlight that falls on them—anywhere from 1% or 2% to almost 40%—into electricity. Solar panels, which are getting cheaper every year, can be installed on the roofs of houses to produce electricity to run refrigerators, computers, TVs, lights, and other machines. The amount of electricity produced by a collection of solar panels depends on their area: the more area, the more electricity. Therefore, whether a system of solar panels can meet all the electricity demands of a household depends on three things: (1) how much electricity the household uses, (2) how efficient the solar panels are (that is, how much of the sun energy that falls on them is turned into electricity), and (3) how much area is available on the roof of the house.

The average U.S. household uses about 9,000 kWh of electricity per year. A kWh, or kilowatt-hour, is the amount of electricity used by a 100-watt light bulb burning for 10 hours. That's equal to 1,040 watts of around-the-clock use, which is the amount of electricity used by ten 100-watt bulbs burning constantly. A typical square meter of land in the United States receives from the Sun about 150 watts of power per square meter (W/m^2), averaged around the clock, so using solar panels with an efficiency of 20% we could harvest about 30 watts per square meter of panel (on average, around the clock). To get 1,040 watts, therefore, we need 1,040 W / 30 W/m^2 = 34 m^2 of solar panels. At a more realistic 10% panel efficiency, we would need twice as much panel area, about 68 m^2. This would be a square 8.2 meters on a side (27 feet). Many household rooftops in the United States could accommodate a solar system of this size, but it would be a tight fit. In Europe and Japan, where the average household uses about half as much electricity as the average U.S. household, it would be easier to meet all of a household's electricity demands using a solar panel system. Of course, it might still a good idea to meet some of a household's electricity needs using solar panels, even where it is not practical to meet them completely that way.

Where to Learn More

Web sites

Math.com. "Area Formulas." 2005. <http://www.math.com/tables/geometry/areas.htm> (March 9, 2005).

Math.com. "Area of Polygons and Circles." 2005. <http://www.math.com/school/subject3/lessons/S3U2L4GL.html> (March 9, 2005).

O'Connor, J.J., E.F. Robertson. "An Overview of Egyptian Mathematics." December 2000. <http://www-groups.dcs.st-and.ac.uk/~history/HistTopics/Egyptian_mathematics.html> (March 9, 2005).

O'Neill, Dennis. "Adapting to Climate Extremes." <http://anthro.palomar.edu/adapt/adapt_2.htm> (March 9, 2005).

Average

Overview

An average is a number that expresses the central tendency of a group of numbers. Another word for average, one that is used more often in science and math, is "mean." Averages are often used when people need to understand groups of numbers. Whenever groups of measurements are collected in biology, physics, engineering, astronomy or any other science, averages are calculated. Averages also appear in grading, sports, business, politics, insurance, and other aspects of daily life. An average or mean can be calculated for any list of two or more numbers by adding up the list and dividing by how many numbers are on it.

Fundamental Mathematical Concepts and Terms

ARITHMETIC MEAN

There are several ways to get at the "average" value of a set of numbers. The most common is to calculate the arithmetic mean, usually referred to simply as "the mean." Imagine any group of numbers—say, 140, 141, 156, 169, and 170. These might stand for the heights in centimeters (cm) of five students. To find their mean, add them up and divide by the number of numbers in the list, in this case, 5:

$$\text{Mean} = \frac{140 + 141 + 156 + 169 + 170}{5}$$

$$= \frac{776}{5}$$

$$= 155.2$$

Figure 1: Calculation of an average or mean.

The average or mean height of the students is therefore 155.2 centimeters (about 5 ft 1 in). Mentioning the mean is a quicker, easier way of describing about how tall the students in the group are than listing all five individual heights.

This is convenient, but to pay for this convenience, information must be left out. The mean is a single number formed by blending all the numbers on the original list together, and can only tell us so much. From the mean, we cannot tell how tall the tallest person or shortest person in the group is, or how close people in the

group tend to be to the mean, or even how big the group is—all things that we might want to know. These details are often given by listing other numbers as well as the mean, such as the minimum (smallest number), maximum (largest number), and standard deviation (a measure of how spread out the list is).

More than one list of numbers might have the same mean. For example, the mean of the three numbers 155, 155.2, and 155.4 is also 155.2.

GEOMETRIC MEAN

The kind of average found by adding up a list of numbers and dividing by how many there are is called the "arithmetic" mean to distinguish it from the "geometric" mean. When numbers on a list are multiplied by each other, they yield a product; the geometric mean of the list is the number that, when multiplied by itself as many times as there are numbers on the list, gives the same product. Take, for example, the list 2, 6, 12. The product of these three numbers is $2 \times 6 \times 12 = 144$. The geometric mean of 2, 6, and 12 is therefore 5.24148 because $5.24148 \times 5.24148 \times 5.24148$ also equals 144.

The geometric mean is not found by adding up the numbers on the list and dividing by how many there are, but by multiplying the numbers together and finding the nth root of the product, where n stands for how many numbers there are on the list. So, for instance, the geometric mean of 2, 6, and 12 is the third (or "cube") root of $2 \times 6 \times 12$:

$$\text{Geometric mean} = \sqrt[3]{2 \times 6 \times 12}$$

$$= \sqrt[3]{144}$$

$$= 5.24148$$

The geometric mean is used much less often than the arithmetic mean. The word "mean" is always taken as referring to the arithmetic mean unless stated otherwise.

THE MEDIAN

Another number that expresses the "average" of a group of numbers is the median. If a group of numbers is listed in numerical order, that is, from smallest to largest, then the median is the number in the middle of the list. For the list 140, 141, 156, 169, 170, the median is 156.

The mean and the median are similar in that they both give a number "in the middle." The difference is that the mean is the "middle" of where the listed numbers are on the number line, whereas the median is just the number that happens to be in the middle of the list. Consider the list 1, 1, 1, 1, 100. The mean is found by adding them up and dividing by how many there are:

$$\frac{1 + 1 + 1 + 1 + 100}{5} = \frac{104}{5} = 20.8$$

The median, on the other hand—the number in the middle of the list—is simply 1. For this particular list, therefore, the mean and median are quite different. Yet for the list of heights discussed earlier (140, 141, 156, 169, 170), the mean is 155.2 and the median is 156, which are similar. What makes the two lists different is that on the list 1, 1, 1, 1, 100, the number 100 is much larger the others: it makes the mean larger without changing the median. (If it were 1 or 10 instead of 100, the median would still be 1—but the average would be smaller.) A number that is much smaller or larger than most of the others on a list is called an "outlier." The rule for finding the median ignores outliers, but the rule for finding the mean does not.

If a list contains an odd number of numbers, as does the five-number list 1, 1, 1, 1, 100, one of the numbers is in the middle: that number is the median. If a list contains an even number of numbers, then the median is the number that lies halfway between the two numbers nearest the middle of the list: so, for the four-number list 1, 1, 2, 100 the median is 1.5 (halfway between 1 and 2).

WHAT THE MEAN MEANS

The mean is not a physical entity. It is a mathematical tool for making sense of a group of numbers. In a group of students with heights 140, 141, 156, 169, 170 cm and average height 155.2 cm, no single person is actually 155.2 cm tall. It does not usually mean much, therefore, when we are told that somebody or something is above or below average. In this group of students, everybody is above or below average.

Further, averages only make sense for groups of numbers that have a gist or central tendency, that are fairly evenly scattered around some central value. Averages do not make sense for groups of numbers that cluster around two or more values. If a room contains a mouse weighing 50 grams and an elephant weighing 1,000,000 grams, you could truly say that the room contains a population of animals weighing, on average, $(50 + 1{,}000{,}0000) / 2 = 500{,}025$ grams, half as much as a full-grown elephant, but

this would be somewhat ridiculous. It is more reasonable to say simply that the room contains a 50-gram mouse and a 1,000,000-gram elephant and forget about averaging altogether in this case. If the room contains a thousand mice and a thousand elephants, it might be useful to talk about the mean weight of the mice and the mean weight of the elephants, but it would still probably not make sense to average the mice and the elephants together. The weights of the mice and elephants belong on different lists because mice and elephants are such different creatures. These two lists will have different means. In general, the average or arithmetic mean of a list of numbers is meaningful only if all the numbers belong on that list.

A Brief History of Discovery and Development

The concept of the average or mean first appeared in ancient times in problems of estimation. When making an estimate, we seek an approximate figure for some number of objects that cannot be counted directly: the number of leaves on a tree, soldiers in an attacking army, galaxies in the universe, jellybeans in a jar. A realistic way to get such a figure—sometimes the only realistic way—is to pick a typical part of the larger whole, then count how many leaves, soldiers, galaxies, or jellybeans appear in that fragment, then multiply this figure by the number of times that the part fits into the whole. This gives an estimate for the total number. If there are 100 leaves on a typical branch, for instance, then we can estimate that on a tree with 1,000 branches there will be 100,000 leaves. By a "typical" branch, we really mean a branch with a number of leaves on it equal to the average or mean number of leaves per branch. The idea of the average is therefore embedded in the idea of estimation from typical parts. The ancient king Rituparna, as described in Hindu texts at least 3,000 years old, estimated the number of leaves on a tree in just this way. This shows that an intuitive grasp of averages existed at least that long ago.

By 2,500 years ago, the Greeks, too, understood estimation using averages. They had also discovered the idea of the arithmetic mean, possibly to help in spreading out losses when a ship full of goods sank. By 300 B.C., the Greeks had discovered not only the arithmetic mean but the geometric mean, the median, and at least nine other forms of average value. Yet they understood these averages only for cases involving two numbers. For example, the philosopher Aristotle (384–322 B.C.) understood that the arithmetic mean of 2 and 10 was 6 (because 2 plus 10 divided by 2 equals 6), but could not have calculated the average height of the five students in the example used earlier. It was not until the 1500s that mathematicians realized that the arithmetic mean could be calculated for lists of three or more numbers. This important fact was discovered by astronomers who realized that they could make several measurements of a star's position, with each individual measurement suffering from some unknown, ever-changing error, and then average the measurements to make the errors cancel out. From the late 1500s on, averaging to reduce measurement error spread to other fields of study from astronomy. By the nineteenth century averaging was being used widely in business, insurance, and finance. Today it is still used for all these purposes and more, including the calculation of grade-point averages in schools.

Real-life Applications

BATTING AVERAGES

A batting average is a three-digit number that tells how often a baseball player has managed to hit the ball during a game, season, or career. A player's batting average is calculated by dividing the number of hits the player gets by the number of times they have been at bat (although this is not the number of times they have stepped up to the plate to hit because there are also special rules as to what constitutes a legal "at bat" to be used in calculating a player's batting average). Say a player goes to bat 3 times and gets 0 hits the first time, 1 the second, and 0 the third (this is actually pretty good). Their batting average is then $(0 + 1 + 0) / 3 = .333$. (A batting average is always rounded off to three decimal places.) A batting average cannot be higher than 1, because a player's turn at bat is over once they get a hit: if a player went up three times and got three hits, their batting average would $(1 + 1 + 1) / 3 = 1.000$.

But this would be superhumanly high. Not even the greatest hitters in the Baseball Hall of Fame got a hit every time they went to bat—or even half the time they went to bat. Ty Cobb, for instance, got 4,191 hits in 11,429 turns at bat for a batting average of .367, the highest career batting average ever. The highest batting average for a single season, .485, was achieved by Tip O'Neill in 1887.

In cricket, popular in much of the world outside the United States, a batsman's batting average is determined by the number of runs they have scored divided by the number of times they have been out. A "bowling average" is calculated for bowlers (the cricket equivalent of pitchers) as the number of runs scored against the bowler divided by the number of wickets they have taken. The

A motorcyclist soars high during motocross freestyle practice at the 2000 X Games in San Francisco. Riders and coaches make calculations of average "hang time" and length of jumps at various speeds so that they know what tricks are safe to land. AP/WIDE WORLD PHOTOS. REPRODUCED BY PERMISSION.

higher a cricket player's batting average, the better; the lower a player's bowling average, the better.

GRADES

In school, averages are an everyday fact of life: an English or algebra grade for the marking period is calculated as an average of all the students' test scores. For example, if you do four assignments in the course of the marking period for a certain class and get the scores 95, 87, 82, and 91, then your grade for the marking period is

$$\frac{95 + 87 + 82 + 91}{4} = 88.75$$

In many schools that assign letter grades, all grades between 80 and 90 are considered Bs. In such a school, your grade for the marking period in this case would be a B.

WEIGHTED AVERAGES IN GRADING

What if some of the assignments in a course are more important than the others? It would not be fair to count them all the same when averaging scores to calculate your grade from the marking period, would it? To make score-averaging meaningful when not all scores stand for equally important work, teachers use the weighted-average method. Calculation of a weighted average assigns a weight or multiplying factor to each grade. For example, quizzes might be assigned a weight of 1 and tests a weight of 2 to signify that they are twice as important (in this particular class). The weighted average is then calculated as the sum of the grades—each grade multiplied by its weight—divided by the sum of the weights. So if during a marking period you take two quizzes (grades 82 and 87) and two tests (grades 95 and 91), your grade for the marking period will be

$$\frac{82 + 87 + (2 \times 95) + (2 \times 91)}{1 + 1 + 2 + 2} = \frac{541}{6} = 90.2$$

Because you did better on the tests than on the quizzes, and the tests are weighted more heavily than the quizzes, your grade is higher than if all the scores had been worth the same.

In most colleges and some high schools, weighted averaging is also used to assign a single number to academic performance, the famous (or perhaps infamous) grade point average, or GPA. Like individual tests, some classes require more work and must be given a heavier weight when calculating the GPA.

WEIGHTED AVERAGES IN BUSINESS

Weighted averages are also used in business. If in the course of a month a store sells different amounts of five kinds of cheese, some more expensive than others, the owner can use weighted averaging to calculate the average income per pound of cheese sold. Here the "weight" assigned to the sales figure for each kind of cheese is the price per pound of that cheese: more expensive cheeses are weighted more heavily. Weighted averaging is also used to calculate how expensive it is to borrow capital (money for doing business) from various lenders that all charge different interest rates: a higher interest rate means that the borrower has to pay more for each dollar borrowed, so money from a higher-interest-rate source costs more. When a business wants to know what an average dollar of capital costs, it calculates a weighted average of borrowing costs. This commonly calculated figure is known in business as the weighted average cost of capital. Spreadsheet software packages sold to businesses for calculating

profit and loss routinely include a weighted-averaging option.

AVERAGING FOR ACCURACY

How long does it take a rat to get sick after eating a gram of Chemical X? Exactly how bright is Star Y? Each rat and each photograph of a star is a little different from every other, so there is no final answer to either of these questions, or to any other question of measurement in science. But by performing experiments on more than one rat (or taking more than one picture of a star, or taking any other measurement more than once) and averaging the results, scientists can get a better answer than if they look at just one measurement. This is done constantly in all kinds of science. In medical research, for instance, nobody performs an experiment or gathers data on just one patient. An observation is performed as many times as is practical, and the measurements are averaged to get a more accurate result. It is also standard practice to look at how much the measurements tend to spread out around the average value—the "standard deviation."

How does averaging increase accuracy? Imagine weighing a restless cat. You weigh the cat four times, but because it won't hold still you get a scale reading each time that is a little too high or a little too low: 5.103 lb, 5.093 lb, 5.101 lb, 5.099 lb. In this case, the cat's real weight is 5.1 lb. The error in the first reading, therefore, is .003 lb, because 5.1 + .003 = 5.103. Likewise, the other three errors are −.003, .001, and −.001 lb. The average of these errors is 0:

$$\frac{.003 + (-.003) + .001 + (-.001)}{4} = \frac{0}{4} = 0$$

The average of the four weights is therefore the true weight of the cat:

$$\frac{5.103 + 5.093 + 5.101 + 5.099}{4} = \frac{20.4}{4} = 5.1$$

Although in real life the errors rarely cancel out to exactly zero, the average error is usually much smaller than any of the individual errors. Whenever measurement errors are equally likely to be positive and negative, averaging improves accuracy.

In astronomy, this principle has been used for the star pictures taken by the International Ultraviolet Explorer satellite, which took pictures of stars from 1978 to 1996. To make final images for a standard star atlas (a collection of images of the whole sky), two or three images for each star were combined by averaging. In fact, a weighted average was calculated, with each image being weighted by its exposure: short-exposure images were dimmer, and were given a heavier weight to compensate. The resulting star atlas is more accurate than it would have been without averaging.

HOW MANY GALAXIES?

As scientists discovered in the early twentieth century, the Universe does not go on forever. It is finite in size, like a very large room (only without walls, and other strange properties). There cannot, therefore, be an infinite number of galaxies because there is not an infinite space.

Scientists use averages to estimate such large numbers. Galaxies, like leaves on a large tree, are hard to count. Many galaxies are so faint and far away that even the powerful Hubble Space Telescope must gaze for days a small patch of sky to see them. It would take many years to examine the whole sky this way, so instead the Hubble takes a picture of just one part of the sky—an area about as big as a dime 75 ft (22.86 m) away. Scientists assume that the number of galaxies in this small area of the sky is about the same as in any other area of the same size. That is, they assume that the number of galaxies in the observed area is equal to the average for all areas of the same size. By counting the number of galaxies in that small area and multiplying to account for the size of the whole sky, they can estimate the number of galaxies in the Universe.

In 2004, the Hubble took a picture called the Ultra Deep Field, gazing for 300 straight hours at one six-millionth of the sky. The Ultra Deep Field found over 10,000 galaxies in that tiny area. If this is a fair average for any equal-sized part of the sky, then there are at least twenty billion galaxies in the universe. Most galaxies contain several hundred billion stars.

THE "AVERAGE" FAMILY

Any list of numbers has an average, but an average that has been calculated for a list of numbers that does not cluster around a central value can be meaningless or misleading. In such a case, the "distribution" of the numbers—how they are clumped or spread out on the number line—can be important. This knowledge is lost when the numbers are squashed down into a single number, the average.

In politics, numbers about income, taxes, spending, and debt are often named. It is sometimes necessary to talk about averages when talking about these numbers, but some averages are misleading. Sometimes politicians, financial experts, and columnists quote averages in a way that creates a false impression.

For example, public figures often talk about what a proposed law will give to or take away from an "average" family. If the subject is income, then most listeners probably assume that an "average" family is a family with an income near the median of the income range. For instance, if 99 families in a certain neighborhood make $30,000 a year and one family makes $3,000,000, the median income will be $30,000 but the average income—the total income of the neighborhood divided by the number of families living there—will be $59,700, twice as much as all but one of the families actually make. To say that the "average" family makes almost $60,000 in this neighborhood would be mathematically correct but misleading to a typical listener. It would make it sound like a wealthier neighborhood than it really is.

This problem is that there is an unusually large value in the list of incomes, namely, the single $3,000,000 income—an outlier. This makes the arithmetic average inappropriate. A similar problem often arises in real life when political claims are being made about tax cuts. A tax cut that gives a great deal of money to the richest one percent of families, and a great deal less money to all the rest, might give an "average" of, say, $2,500.00 each year. "My tax cut will put $2,500 back in the pocket of the average American family!" a politician might say, meaning that the sum of all tax cuts divided by the number of all families receiving cuts equals $2,500.00. Yet only a small number of wealthier families might actually see cuts of $2,500 or larger. Middle-class and poorer families, to whom the number "$2500.00" sounds more important because it a bigger percentage of their income—the great majority of voters hearing the politician's promise—might actually have no chance of receiving as much as $2,500. An average figure can misused to convey a false idea while still being mathematically true.

SPACE SHUTTLE SAFETY

Many of the machines on which lives depend—jet planes, medical devices, spacecraft, and others—contain thousands or millions of parts. No single part is perfectly reliable, but in designing complex machines we would like to guarantee that the chances of a do-or-die part failing during use is very small. But how do we put a number on a part's chances for failing?

For commonplace parts, one way is to hook up a large number of them and watch to see how many fail, on average, in a given period of time. But for a complex system like a space shuttle, designers cannot afford to wait and they cannot afford to fail. They therefore resort to a method known as "probabilistic risk assessment." Probabilistic risk assessment tries to guess the chances of the complex system failing based on the reliability of all its separate parts. Reliability is sometimes expressed as an average number, the "mean time between failures" (MBTF). If the MBTF for a computer hard drive is five years, for example, then after each failure you will have to wait—on average—five years until another failure occurs. The MBTF is not a minimum, but an average: the next failure might happen the next day, or not for a decade.

MBTF is not an average from real data, but a guess about the average value of numbers that one does not know yet. MBTF estimates can, therefore, be wrong. In the 1980s, in the early days of the space shuttle program, NASA calculated an estimated MBTF for the space shuttle. Its estimate was that the shuttle would suffer a catastrophic accident, on average, during 1 in every 100,000 launches. That is, the official MBTF for the shuttle was 100,000 launches.

But it was at the 25th shuttle launch, that of the space shuttle *Challenger*, that a fatal failure occurred. Seventy-six seconds after liftoff, *Challenger* exploded. This did not prove absolutely that the MBTF was wrong, because the MBTF is an average, not a minimum—yet the chances were small that an accident would have happened so soon if the MBTF were really 100,000 launches. NASA therefore revised its MBTF estimate down to 265 launches. But in 2003, only 88 flights after the *Challenger* disaster, *Columbia* disintegrated during re-entry into the atmosphere. Again, this did not prove that NASA's MBTF was wrong, but if it were right then such a quick failure was very unlikely.

STUDENT LOAN CONSOLIDATION

Millions of students end up owing tens of thousands of dollars in student loans by the time they finish college. Usually this money is borrowed in the form of several different loans having different interest rates. After graduation, many people "consolidate" these loans. That is, several loans are combined into one loan with a new interest rate, and this new, single loan is owed to a different institution (usually one that specializes in consolidated loans). There are several advantages to consolidation. The new interest rate is fixed, that is, it cannot go up over time. Also, monthly payments are usually lower, and there is only one payment to make, rather than several.

The interest rate on a consolidated student loan is calculated by averaging the interest rates for all the old loans that are being consolidated. Say you are paying off two (rather small) student loans. You still owe $100 on one loan at 7% interest and $200 on another at 8% interest. When the loans are consolidated you will owe

$100 + $200 = $300, and the interest rate will be the weighted average of the two interest rates:

$$\text{New interest rate} = \frac{100 \times .07 + 200 \times .08}{300} = .07667$$

The weights in the weighted average are the amounts of money still owed on each loan: the interest rate of the bigger loan counts for more in calculating the new interest rate, which is 7.667%. In practice, the rate is rounded up to the nearest one eighth of a percent, so your real rate would be 7.75%.

AVERAGE LIFESPAN

We often read that the average human lifespan is increasing. Strictly speaking, this is true. In the mid nineteenth century, the average lifespan for a person in the rich countries was about 40 years; today, thanks to medical science and public health advances such as clean drinking water, it is about 75 years. Here the word "average" means the arithmetic mean, that is, the sum of all individual lifespans in a certain historical period divided by the number of people born in that period.

Some have argued that because average lifespan has been increasing, it must keep on increasing without limit, making us immortal. For example, computer scientist Ray Kurzweil said in "the eighteenth century, we added a few days to the human life expectancy every year. In the nineteenth century, we added a few weeks every year. Now we're adding over a hundred days per year to human life expectancy . . . Many observers, including myself, believe that within ten years we will be adding more than a year—every year—to human life expectancy. So as you go forward a year, human life expectancy will move away from us." (Kurzweil, R. "The Ascendence of Science and Technology [a panel discussion]." *Partisan Review*. Sept 2, 2002.)

The problem with this argument is that it mixes up average lifespan with maximum lifespan. The average lifespan is not increasing because people are living to be older than anyone ever could in the past: they are not. A few people have always lived to be 90, 100, or 110 years old. The reason average lifespan is higher now than in the past is that fewer people are dying in childhood and youth. Today, at least in the industrialized countries, most people do not die until old age. However, the ultimate limit on how old a person can get has not increased, and the average lifespan cannot be increased beyond that limit by advances that keep people from dying until they reach it. Perhaps in the future, medical science will increase the maximum possible age, but that is only a possibility. It has nothing to do with past increases in average lifespan.

INSURANCE

In the industrial world, virtually everyone, from their late teens on up, has some kind of insurance. For example, all European Union states and most U.S. states require that all drivers buy liability insurance—that is, insurance to pay for medical care for anyone that the driver may injure in an accident that is their fault. Insurance is basic to business, health care, and personal life—and it is founded on averages.

Insurance companies charge their customers a certain amount every month, a "premium," in return for a commitment that the insurance company will pay the customer a much greater amount of money if a problem should happen—sickness, car accident, death in the family, house fire, or other (depending on the kind of insurance policy). This premium is based on averages. The insurance company groups people (on paper) by age, gender, health, and other factors. It then calculates what the average rate of car wrecks, house fires, or other problems for the people in each group, and how much these problems cost on average. This tells it how much it has to charge each customer in order to pay for the money that the company will have to pay out—again, on average. To this amount is added the insurance company's cost of doing business and a profit margin (if the insurance company is for-profit, which not all are).

Insurance costs are higher for some groups than for others because they have higher average rates for some problems. For example, young drivers pay more for car insurance because they have more accidents. The average crash rate per mile driven for 16-year-olds is three times higher than for 18- and 19-year olds; the rate for drivers 16–19 years old, considered as a single group, is four times higher than for all older drivers. What's more, young male drivers 16–25, who on average drive more miles, drink more alcohol, and take more driving risks, have more accidents than female drivers in this age group: two thirds of all teenagers killed in car crashes (the leading cause of death for both genders in the 18–25 age group) are male.

More crashes, injuries, and deaths mean more payout by the insurance company, which makes it reasonable, unfortunately, for the company to charge higher rates to drivers in this group. Some companies offer reduced-rate deals to young drivers who avoid traffic tickets.

EVOLUTION IN ACTION

Averaging makes it possible to see trends in nature that can't be seen by looking at individual animals. Averages have been especially useful in studying evolution, which happens to slowly to see by looking at individual

animals and their offspring. The most famous example of observed evolutionary changes is the research done by the biologists Peter and Rosemary Grant on the Galapagos Islands off the west coast of South America. Fourteen or 15 closely related species of finches live in the Galapagos. The Grants have been watching these finches carefully for decades, taking exact measurements of their beaks. They average these measurements together because they are interested in how each finch population as a whole is evolving, rather than in how the individual birds differ from each other. The individual differences, like random measurement errors, tend to cancel each other out when the beak measurements are averaged. When a list of data is averaged like this, the resulting mean is called a "sample mean."

The Grants' measurements show that the average beak for each finch species changes shape depending on what kind of food the finches can get. When mostly large, tough seeds are available, birds with large, seed-cracking beaks get more food and leave more offspring. The next generation of birds has, on average, larger, tougher beaks. This is exactly what the Darwinian theory of evolution predicts: slight, inherited differences between individual animals enable them to take advantage of changing conditions, like food supply. Those birds whose beaks just happen to be better suited to the food supply leave more offspring, and future generations become more like those successful birds.

Key Terms

Mean: Any measure of the central tendency of a group of numbers.

Median: When arranging numbers in order of ascending size, the median is the value in the middle of the list.

Where to Learn More

Books

Tanur, Judith M., et al. *Statistics: A Guide to the Unknown.* Belmont, CA: Wadsworth Publishing Co., 1989.

Wheater, C. Philip, and Penny A. Cook. *Using Statistics to Understand the Environment.* New York: Routledge, 2000.

Web sites

Insurance Institute for Highway Safety. "Q7&A: Teenagers: General." March 9, 2004. <http://www.iihs.org/safety_facts/qanda/teens.htm#2> (February 15, 2005).

Mathworld. "Arithmetic mean." Wolfram Research. 1999. <http://mathworld.wolfram.com/ArithmeticMean.html> (February 15, 2005).

Wikelsky, Martin. "Natural Selection and Darwin's Finches." Pearson Education. 2003. <http://wps.prenhall.com/esm_ freeman_evol_3/0,8018,849374-,00.html> (February 15, 2005).

Base

Overview

In everyday life, a base is something that provides support. A house would crumble if not for the support of its base. So it is too with math. Various bases are the foundation of the various ways we humans have devised to count things. Counting things (enumeration) is an essential part of our everyday life. Enumeration would be impossible if not for based valued numbers.

Fundamental Mathematical Concepts and Terms

In numbering systems, the base is the positive integer that is equal to the value of 1 in the second highest counting place or column. For example, in base 10, the value of a 1 in the "tens" column or place is 10.

A Brief History of Discovery and Development

The various base numbering systems that have arisen since before recorded history have been vital to our existence and have been one of the keys that drove the formation of societies. Without the ability to quantify information, much of our everyday world would simply be unmanageable. Base numbering systems are indeed an important facet of real life math.

The concept of the base has been part of mathematics since primitive humans began counting. For example, animal bones that are about 37,000 years old have been found in Africa. That is not the remarkable thing. The remarkable thing is that the bones have human-made notches on them. Scientists argue that each notch represented a night when the moon was visible. This base 1 (1, 2, 3, 4, 5, . . .) system allowed the cave dwellers to chart the moon's appearance. So, the bones were a sort of calendar or record of the how frequent the nights were moonlit. This knowledge may have been important in determining when the best was to hunt (sneaking up on game under a full moon is less successful than when there is no moon).

Another base system that is rooted in the deep past is base 5. Most of us are familiar with base 5 when we chart numbers on paper, a whiteboard or even in the dirt, by making four vertical marks and then a diagonal line across these. The base 5-tally system likely arose because of the construction of our hands. Typically, a hand has four fingers and a thumb. It is our own carry-around base 5 counting system.

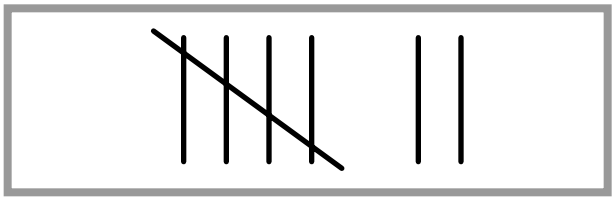


Figure 1: Counting to seven in a base 5 tally system.

In base 5 tallying, the number 7 would be represented as depicted in Figure 1.

Of course, since typically we have two hands and a total of ten digits, we can also count in multiples of 10. So, most of us also naturally carry around with us a convenient base 10 (or decimal) counting system.

Counting in multiples of 5 and 10 has been common for thousands of years. Examples can be found in the hieroglyphics that adorn the walls of structures built by Egyptians before the time of Christ. In their system, the powers of 10 (ones, tens, hundreds, thousands, and so on) were represented by different symbols. One thousand might be a frog, one hundred a line, ten a flower and one a circle. So, the number 5,473 would be a hieroglyphic that, from left to right, would be a pattern of five frogs, four lines, seven flowers and three circles.

There are many other base systems. Base 2 or binary (which we will talk about in more detail in the next section) is at the heart of modern computer languages and applications. Numbering in terms of groups of 8 is a base-8 (octal) system. Base 8 is also very important in computer languages and programming. Others include base 12 (duodecimal), base 16 (hexidecimal), base 20 (vigesimal) and base 60 (sexagesimal).

The latter system is also very old, evidence shows its presence in ancient Babylon. Whether the Babylonians created this numbering system outright, or modified it from earlier civilizations is not clear. As well, it is unclear why a base 60 system ever came about. It seems like a cumbersome system, as compared with the base 5 and 10 systems that could literally rely on the fingers and some scratches in the dirt to keep track of really big numbers. Even a base 20 system could be done manually, using both fingers and toes.

Scholars have tried to unravel the mystery of base 60's origin. Theories include a relationship between numbers and geometry, astronomical events and the system of weights and measures that was used at the time. The real explanation is likely lost in the mists of time.

Real-life Applications

BASE 2 AND COMPUTERS

Base 2 is a two digit numbering system. The two digits are 0 and 1. Each of these is used alternately as numbers grow from ones to tens to hundreds to thousands and upwards. Put another way, the base 2 pattern looks like this: 0, 1, 10, 11, 100, 101, 110, 111, 1000, . . . (0, 1, 2, 3, 4, 5, 6, 7, 8, . . .).

The roots of base 2 are thought to go back to ancient China but base 2 is as also fresh and relevant because it is perfect for the expression of information in computer languages. This is because, for all their sophistication, computer language is pretty rudimentary. Being driven by electricity, language is either happening as electricity flows (on) or it is not (off). In the binary world of a computer, on is represented by 1 and off is represented by 0.

As an example, consider the sequence depicted in Figure 2.

off-off-on-off-on-on-on-off-on-on

Figure 2: Information series.

In the base 2 world, this sequence would be written as depicted in Figure 3.

0010111011

Figure 3: Information series translated to Base 2.

View the fundamental code for a computer program and you will see line upon line of 0s and 1s. Base 2 in action!

Each 0 or 1 is known as a bit of information. An arrangement of four bits is called a nibble and an arrangement of 8 bits is called a byte (more on this arrangement below, in the section on base 8).

A base 2 numbering system can also involve digits other than 0 and 1, with the arrangement of the numbers being the important facet. In this arrangement, each number is double the preceding number. This base 2 pattern looks like this: 1, 2, 4, 8, 16, 32, 64, 128, 256, . . . It is also evident that in this series, from one number to the next, the numbers of the power also double. For example, compare the numbers 64 and 128. In the larger number, 12 is the double of 6 and 8 is the double of 4.

Base 8

In the base 8 number system, each digit occupies a place value (ones, eights, sixteens, etc.). When the number 7 is reached, the digit in that place switches back to 0 and 1 is added to the next place. The pattern looks like this: 0, 1, 2, 3, 4, 5, 6, 7, 10, 11, 12, 13, 14, 15, 16, 17, 20, 21, 22, ...

Each increasing place value is 8 times as big as the preceding place value. This is similar to the pattern shown above for base 2, only now the numbers get a lot bigger more quickly. The pattern looks like this: 1, 8, 64, 512, 4096, 32768, . . .

As mentioned in the preceding section, the base 2 digits can be arranged in groups of 8. In the computer world, this arrangement is called a byte. Often, computer software programs are spoken of in terms of how many bytes of information they consist of. So, the use of the base 8 numbering system is vital to the operation of computers.

Base 10

The base 10, or decimal, numbering system is another ancient system. Historians think that base 10 originated in India some 5,000 years ago.

The digits used in the base 10 system are 0 through 9. When the latter is reached, the value goes to 0 and 1 is added to the next place. The pattern look like this: 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, . . .

Each successive place value is 10 times greater than the preceding value, which results in the familiar ones, tens, hundreds, thousands, etc. columns with which we usually do addition, subtraction, multiplication and division.

Where to Learn More

Books

Devlin, K.J. *The Math Gene: How Mathematical Thinking Evolved & Why Numbers are like Gossip*. New York: Basic Books, 2001.

Gibilisco, S. *Everyday Math Demystified*. New York: McGraw-Hill Professional, 2004.

Web sites

Loy, J. "Base 2 (Binary)." <http://www.jimloy.com/math/base2.htm> (October 31, 2004).

Poseidon Software and Invention. "Base Valued Numbers." <http://www.psinvention.com/zoetic/basenumb.htm> (October 31, 2004).

Smith, J. "Base Arithmetic." <http://www.jegsworks.com/Lessons/reference/basearith.htm> (October 30, 2004).

Business Math

Overview

Money is the difference between leisure activity and business. While enjoying leisure activity one can expect to pay to have a good time by purchasing a ticket, supplies or paying a fee to gain access to whatever they wish to do. Business activity in any form spends money to earn money. In both cases, numbers are the alphabet of money and math is its universal language.

Computing systems have displaced manual information gathering, recordkeeping, and accounting at an ever-increasing rate within the business world. Advancing computer technology has made this possible and, to some extent, decreasing math skills among the general populations of all nations have made it necessary. One of the initial motivating factors that have led more and more stores to investing large amounts of money to install and operate code-scanning checkout systems is the increasing difficulty in finding an adequate number of people with the necessary math skills to consistently and reliably make change at checkout counters. The introduction of these systems has improved merchants' ability to keep accurate records of what they sell, what they need to order, and to recognize what their customers want so that they may maintain a ready supply. However, for all of the advances business computing has made in generating real-time management reports, none of it is of any value without people who can interpret what it means and, to do that, one must understand the math used by the computing system. Simply because a computer prints out a report does not ensure that it is accurate or useful.

It is worth stating that those people with good math skills will have the best opportunities to excel in many ways in jobs and careers within the business world. Math is not just an exercise for the classroom, but is a critical skill if one is to succeed now and in the future. All money is being monitored and managed by someone. One's personal future depends on how well they manage their money. The future of any employer, and the local, state, and national governments in which one lives, depends on how well they manage money. Money attracts attention. If a person or the business and the governmental institutions they depend on do not use the math skills necessary to wisely manage the money in their respective care, someone else will and they are not likely have the best interest of others in mind. Math skills are one of the most essential means for one to look after their own best interest as an individual, employee, investor, or business owner.

Fundamental Mathematical Concepts and Terms

Business math is a very broad subject, but the most fundamental areas include budgets, accounting, payroll, profits and earnings, and interest.

BUDGETS

All successful businesses of any size, from single individuals to world-class corporations, manage everything according to a budget. A budget is a plan that considers the amount of money to be spent over a specific time schedule, what it is to be spent on, how that money is to be obtained, and what it is expected to deliver in return. Though this sounds simple, it is a very complicated concept.

Businesses and governments rise and fall on their ability to perform reliably according to their budgets. Budgets include detailed estimates of money and all related activities in a format that enables the state of progress toward established goals and objectives to be monitored on a regular basis through various business reports. The reports provide the information necessary for management to identify opportunity and areas of concern or changing conditions so that proper adjustments may be made and put into action in timely fashion to improve the likelihood of success or warn of impending failure to meet expectations. In a budget, all actions, events, activities, and project outcomes are quantified in terms of money.

The basic components of any budget are capital investments, operating expense and revenue generation. Capital investments include building offices, plants and factories, and purchasing land or equipment and the related goods and services for new projects, including the cost of acquiring the money to invest in these projects. Expense outlays include personnel wages, personnel benefits, operating goods and services, advertising, rents, royalties, and taxes.

Budgets are prepared by identifying and quantifying the cost and contributions from all ongoing projects, as well as new projects being put in place and potential new projects and opportunities expected to be begun during the planning cycle. Typically, budgets cover both the immediate year and a longer view of the next three to five years. Historical trends are derived by taking an after-look at the actual results of prior period budgets compared to their respective plan projections. Quite often the numerical data is converted to graphs and charts to aid in spotting trends and changes over time. A simple budget is represented by Figure 1.

The math involved in this simplistic example budget is addition, subtraction, and multiplication, where Revenue from shoe and sandal sales = Number of pairs of sold multiplied by the price received; Personnel Expense = Number of people employed each month multiplied by individual monthly wages; Federal Taxes = The applicable published tax rate multiplied times Income Before Tax.

As the year progresses, a second report would be prepared to compare the projections above with the actual performance. If seasonal shoe sales fall below plan, then the company knows that they need to improve the product or find out why it is not selling as expected. If shoe sales are better than expected, they may need to consider building another factory to meet increasing demand or acquire additional shoes elsewhere.

This somewhat boring exercise is essential to the A.Z. Neuman Shoe Factory to know if it is making or losing money and if it is a healthy company or not. This information also helps potential investors decide if the company is worth investing money in to help grow, to possibly buy the company itself, or to sell if they own any part of it. As a single year look at the company, A.Z. Neuman seems to be doing fine. To really know how well the company is doing, one would have to look at similar combined reports over the past history of the company, its outstanding debts, and similar information on its competitors.

ACCOUNTING

Accounting is a method of recordkeeping, commonly referred to as bookkeeping, that maintains a financial record of the business transactions and prepares various statements and reports concerning the assets, liabilities, and operating performance of a business. In the case of the A.Z. Neuman Shoe Factory, transactions include the sale of shoes and sandals, the purchase of supplies, machines, and the building of a new store as shown in the budget. Other transactions not shown in detail in the budget might include the sale of stocks and bonds or loans taken to raise the necessary money to buy the machines or build the new store if the company did not have the money on hand from prior years' profits to do so.

People who perform the work of accounting are called accountants. Their job is to collect the numbers related to every aspect of the business and put them in proper order so that management can review how the company is performing and make necessary adjustments. Accountants usually write narratives or stories that serve to explain the numbers. Computing systems help gather and sort the numbers and information, and it is very important that the accountant understand where the

A. Z. Neuman Shoe Factory - Projected Annual Budget – Figures rounded to $MM (millions)

Months:	**J**	**F**	**M**	**A**	**M**	**J**	**J**	**A**	**S**	**O**	**N**
	D	Total									
Revenue											
Shoe sales		3	4	4	16	14	2	2	3	18	4
	3	1	74								
Sandal sales		0	1	1	3	4	3	3	2	1	1
	0	0	19								
Total		3	5	5	19	18	5	5	5	19	5
	3	1	93								
Operating Expense											
Personnel		1	2	2	2	1	1	1	1	1	1
	1	1	15								
Supplies	2	2	2	2	2	2	2	2	1	1	1
	1	20									
Electricity		1	1	1	1	1	0	1	0	1	0
	1	0	8								
Local Taxes		0	0	0	1	0	0	1	0	0	0
	0	1	3								
Total		4	5	5	6	4	3	5	3	3	2
	3	3	46								
Net Contribution (Revenue – OpExp.)											
		–1	0	0	13	14	2	0	2	16	3
	0	–2	47								
Capital Investments											
Machines		1	3	3	4	0	0	0	0	0	0
	0	0	11								
New Store		0	0	0	0	5	0	0	0	0	0
	0	0	5								
Total		1	3	3	4	5	0	0	0	0	0
	0	0	16								
Income Before Tax (IBT = Net Contribution – Capital)											
		–2	–3	–3	9	9	2	0	2	16	3
	0	–2	31								
State & Federal Tax (Minus = credit)											
		–1	–1	–2	3	3	1	0	0	5	1
	0	–1	8								
Income After Tax (IAT = IBT – S&FT)											
		–1	–2	–1	6	6	1	0	2	11	2
	0	–1	23								

Figure 1: A simple budget.

computing system got its information and what mathematical functions were performed to produce the tables, charts, and figures in order to verify that the information is true and correct. Management must understand the accounting and everything involved in it before it can fully understand how well the company is doing.

When this level of understanding is not achieved for any reason, the performance of the company is not likely

Troy McConnell, founder of Batanga.com at his office. The center broadcasts alternative Hispanic music on dedicated Internet channels to consumers between 12 and 33 years of age. In addition, studies at the center include all aspects of business math. AP/WIDE WORLD PHOTOS. REPRODUCED BY PERMISSION.

to be as expected. It would be like trying to ride a bicycle with blinders on: one hopes to make to the corner without crashing, but odds are they will not. Recent and historical news articles are full of stories of successful companies that achieved positive outcomes because they were aware of what they were doing and managed it well. However, there are almost as many stories of companies that did not do well because they did not understand what they were truly doing and mismanaged themselves or misrepresented their performance to investors and legal authorities. If they only mismanage themselves, companies go out of business and jobs are lost and past investments possibly wasted. If a company misrepresents itself either because it did not keep its records properly, did not do its accounting accurately, or altered the facts and calculations in any untruthful way, people can go to jail. The truth begins with honest mathematics and numbers.

PAYROLL

Payroll is the accounting process of paying employees for the work performed and gathering the information for budget preparation and monitoring. An employee sees how much money is received at the end of a pay period, while the employer sees how much it is spending each pay period and the two perspectives do not see the same number. Why? A.Z. Neuman wants to attract quality employees so it pays competitive wages and provides certain benefits. Tom Smith operates a high-tech machine that is critical to the shoe factory on a regular 40-hour-per-week schedule, has been with the company a few years, and has three dependents to care for. How much money does Tom take home and what does it cost A.Z. Neuman each month? Figure 2 lays out the details.

This is just an example. Not all companies offer such benefits, and the relative split in shared cost may vary considerably if the cost is shared at all. If Tom is a member of a labor union, dues would also be withheld. As is shown in Figure 2, the company has to spend approximately $2 for every $1 Tom takes home as disposable income to live on. Correspondingly, Tom will take home only about half of any raise or bonus he receives from the company. At the end of each tax year, Tom then has to file both State and Federal income tax and may discover that

	Tom	A.Z. Neuman	Government
Gross Pay ($25/hour, 40 hours/week, 4 weeks per month)	$4,000	-$4,000	
Withholdings: (Required by law)			
Social Security 12.4% split 6.2% each	-$248	-$248	$ 496
Medicare 2.9% split 1.45% each	-$58	-$58	$ 116
State & Federal Unemployment Insurance		-$120	$ 120
Federal Income Tax	-$800		$ 800
State Income Tax	-$200		$ 200
Savings Plan (Tom can put up to 4%, company matches)	-$160	-$160	
Insurance (Cost split between Tom and company)			
Life	-$62	-$62	
Medical	-$72	-$72	
Net Pay	$2,400	-$4,720	$1,732

Figure 2: Sample of payroll accounting.

he is either due a refund or owes even more depending on his individual situation. Tax withholdings are required by State and Federal law at least in part to fund the operation of governmental functions throughout the year. In some regions of the country, there are other local and city taxes not shown in this example. If people had to send their tax payments in every month in place of having them automatically withheld they would be more mindful of the burden of taxation. In theory, Tom will get his contributions to Social Security and Medicare back in the future in his old age. Tom's contribution to the savings plan is his own attempt to ensure his future.

PROFITS

Unless Mr. A.Z. Neuman just really enjoyed making shoes, he founded the business to make a profit. A profit is realized when the income received is greater than the sum of all expenditures. As shown in the example budget in Figure 1, the company does not make a profit every month and is very dependent on a few really good months when shoe sales are in season to yield a profit for the year. Most businesses operate in this up and down environment. Some business segments have even longer profit and loss cycles, such that they may lose money for several years before experiencing a strong year and hopefully making enough profit to sustain them through the next down cycle. If they fail to make a profit long enough, companies go out of business and this occurs to a large percentage of all companies every year. Without the effective application of good math skills in accounting and business evaluations as well as the ability to understand their meaning to direct future decisions, companies have no idea if they are in fact growing or dying, but they can be sure they are doing one or the other.

EARNINGS

Fundamentally, profits and earnings are defined in very similar terms. However, earnings are often thought of as the return on capital investments as distinguished from expense, as shown in the budget example in Figure 1. One has to know how much capital has been invested throughout the life of the company to fairly calculate the earnings or return on capital employed. The example budget is limited to only one year and suggests that A.Z. Neuman is expected to earn $23 million while investing $16 million in that year. The budget shows that the company had product to sell before investing in new equipment and a new store; thus, the year shown is benefiting from prior year investments of some unknown magnitude. In the developing period of any company, annual earnings are negative (losses) until the initial investments have generated earnings of equal amount to reach what is called payout. Once past payout, companies can begin generating a positive annual return on capital employed. Some industries require continued annual capital investment to expand or replace their asset base, and this will continue to hold down their annual rate of return until such time as there are no more attractive investment opportunities and they are in the later stage, but high earnings generating phase, of their business life. Typically, businesses that can

A presentation in a "business" environment. PREMIUM STOCK/CORBIS.

generate a 15% rate of return on capital employed over a period of several years have done very well. Most companies struggle to deliver less than half that level of earnings.

INTEREST

Interest is money earned on money loaned or money paid on money borrowed. Interest rates vary based on a variety of factors determined in financial markets and by governmental regulations. Low interest rates are good for a borrower or anyone dependent on others' ability to borrow money to buy goods and services. High interest rates are good for those saving or lending. When the A.Z. Neuman Shoe Factory wants to buy additional equipment or build new factories or stores, it has to determine where the money will come from to do so. If interest rates are low, it may elect to borrow instead of spending its own cash. If interest rates are high, it will have to consider other courses of raising the money needed to fund investments if it has a cash reserve and wishes to hold on to it for protection or other investments. The two primary ways businesses raise capital, other than borrowing, are to sell stocks and bonds in the company.

A share of stock represents a fractional share of ownership in the company for the price paid. The owner of stock shares in the future performance of the company. If the company does well, the stock goes up and the investor does well, and can do very well under the right circumstances. If the company does poorly, the investor does poorly and can lose the entire amount invested. Stock ownership has a definite share of risk while it has a definite attraction of significant growth potential. Companies will pay a return, or dividend, that might be thought of as interest to stockholders when it can afford to do so as incentive for them to continue to own the stock.

Bonds are generally less risky than stocks, but only those ensured by cash reserves or the assets of sound national governments are secure. A company issues a bond, or guaranty, to investors willing to buy them that over a specified period of time interest will be paid on the amount invested and that the original investment will be returned to the buyer when the bond matures. However, the security of a bond is only as good as the company issuing it. It is in the best interest of a company to meet its bond obligations or it may never sell another bond.

Key Terms

Balance: An amount left over, such as the portion of a credit card bill that remains unpaid and is carried over until the following billing period.

Bankruptcy: A legal declaration that one's debts are larger than one's assets; in common language, when one is unable to pay his bills and seeks relief from the legal system.

Interest: Money paid for a loan, or for the privilege of using another's money.

The advantage of a bond to the company is that ownership is not being shared among the buyers, the upside potential of the company remains owned by the company, and the interest rate paid out is usually less than the interest rate that would have to be paid by the company on a loan. The benefit to the buyer is that bonds are not as risky as stock and, while the return is limited by the established interest rate, the initial investment is not at as great a risk of loss. Bonds are safer investments than stocks in that they tend to have guaranteed earnings, even if considerably lower than the growth potential of stock without the downside risk of loss.

Companies pay the interest on loans, the interest on bonds, and any dividends to stockholders out of their earnings; thus, the rate of return as mentioned earlier is an important indicator to potential investors of all types. The assessment of business risks and opportunity can only be performed through extensive mathematical evaluation, and the individuals performing these evaluations and using them to consider investments must possess a high degree of math skills. In the end, the primary difference between evaluating a business and balancing one's own personal checkbook is the magnitude of the numbers.

Where to Learn More

Books

Boyer, Carl B. *A History of Mathematics*. New York: Wiley and Sons, 1991.

Bybee, L. *Math Formulas for Everyday Living*. Uptime Publications, 2002.

Devlin, Keith. *Life by the Numbers*. New York: Wiley and Sons, 1998.

Westbrook, P. *Math Smart for Business: Essentials of Managerial Finance*. Princeton Review, 1997.

Calculator Math

Overview

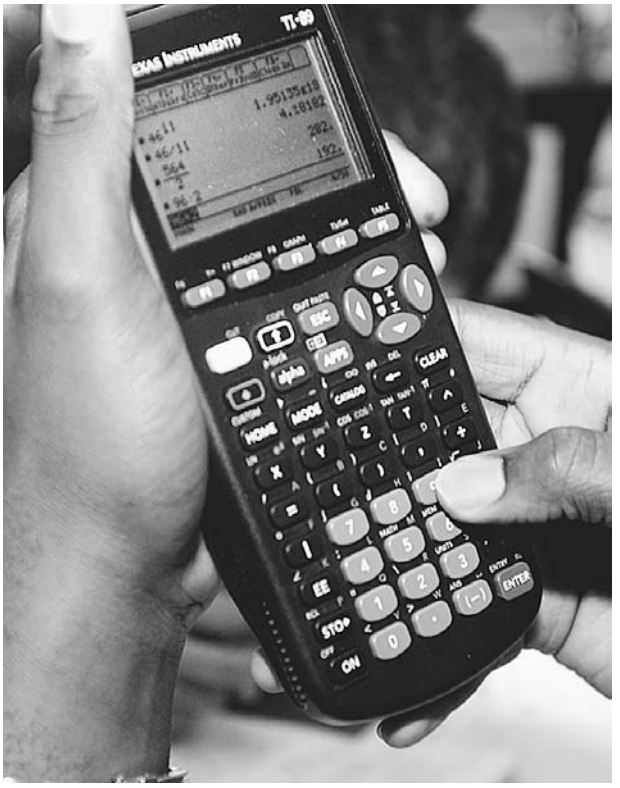
A calculator is a tool that performs mathematical operations on numbers. Some of the simplest calculators can only perform addition, subtraction, multiplication, and division. More sophisticated calculators can find roots, perform exponential and logarithmic operations, and evaluate trigonometric functions in a fraction of a second. Some calculators perform all of these operations using repeated processes of addition.

Basic calculators come in sizes from as small as a credit card to as large as a coffee table. Some specialized calculators involve groups of computing machines that can take up an entire room. A wide variety of calculators around the world perform tasks ranging from adding up bills at retail stores to figuring out the best route when launching satellites into orbit. Calculators, in some form or another, have been important tools for mankind throughout history. Throughout the ages, calculators have progressed from pebbles in sand used for solving basic counting problems to modern digital calculators that come in handy when solving a homework problem or balancing a checkbook.

People regularly use calculators to aid in everyday calculations. Some common types of modern digital calculators include basic calculators (capable of addition, subtraction, multiplication, and division), scientific calculators (for dealing with more advanced mathematics), and graphing calculators. Scientific calculators have more buttons than more basic calculators because they can perform many more types of tasks. Graphing calculators generally have more buttons and larger screens allowing them to display graphs of information provided by the user. In addition to providing a convenient means for working out mathematical problems, calculators also offer one of the best ways to verify work performed by hand.

Fundamental Mathematical Concepts and Terms

Modern calculators generally include buttons, an internal computing mechanism, and a screen. The internal computing mechanism (usually a single chip made of silicon and wires, called a microprocessor, central processing unit, or CPU) provides the brains of the calculator. The microprocessor takes the numbers entered using the buttons, translates them into its own language, computes the answer to the problem, translates the answer back into our numbering system, and displays the answer on the screen. What is even more impressive is that it usually does all of this in a fraction of a second.



A student works on his Texas Instruments graphing calculator. American students have been using graphing calculators for over a decade, and Texas Instruments accounts for more than 80% of those sales, according to an industry research firm. Texas Instruments faces what may turn out to be a more serious challenge: software that turns handheld computers into graphing calculators. AP/WIDE WORLD PHOTOS. REPRODUCED BY PERMISSION.

The easiest way to understand the language of a calculator is to compare it to our numbering system, which is a base ten system. This is due to the fact that we have ten fingers and ten toes. For example, consider how humans count to 34 using fingers. You basically keep track of how many times you count to ten until you get to three, and then count four more fingers. This idea is represented in our numbering system. There is a three in the tens column and a four in the ones column. The tens column represents how many times we have to go through a set of ten fingers, and the ones column represents the rest of the fingers required. A calculator counts in a similar way, but its numbering system is based on the number two instead of ten. This is known as a binary numbering system, meaning that it is based on the number two.

Our ten-based numbering system is known as the decimal numbering system. Much in the same way that each column of a decimal number represents one of the ten numbers between zero and nine, a number in binary form is represented by a series of zeros and ones. Though binary numbers may seem unintuitive and confusing, they are simpler than decimal numbers in many ways, allowing complex calculations to be carried out on tiny microprocessor chips.

The columns (places) in the decimal numbering system each represent multiples of ten: ones, tens, hundreds, thousands, and so forth. After the value of a column reaches nine, the next column is increased. Similarly, the columns in binary numbers represent multiples of two: ones, twos, fours, eights, and so on. Counting from zero, binary numbers go 0, 1, 10, 11, 100, 101, 110, 111, 1000, etc. 110 represents six because it has a one in the fours column, a one in the twos column, and a zero in the ones column. Because binary notation only involves two values in different columns, it is common to think of each column either being on or off. If a column has a 1 in it, then the value represented by the column (1, 2, 4, 8, 16, 32, and so on) is included in the number. So a 1 can be seen to mean that the column is on, and a 0 can be seen to mean that the column is off. This is the essence of the binary numbering system that a calculator uses to perform mathematical operations.

As an example, add the numbers 6 and 7 together. Using fingers to count in decimal numbers, count 6 fingers and then count 7 more fingers. When all ten fingers are used, make a mental tally in the tens column, and then count the last three fingers to get a single tally in the tens column and three in the ones column. This represents one ten and three ones, or 13. When you input 6 plus 7 into a calculator, the calculator firsts translates the two numbers into binary notation. In binary notation, 6 is represented by 110 (a one in the fours column, a one in the twos column, and a zero in the ones column) and 7 is represented by 111 (a one in the fours column, a one in the twos column, and a one in the ones column). Next, the two numbers are added together by adding the columns together. First, adding up the values in the ones column (0 and 1) results in a one in the ones column. Next, adding the values in the twos column results in a 2 so the twos column of the sum get a 0 and the next column over, the fours column, is increased by one (just like the next column in the decimal numbering system is increased when a column goes beyond nine). Adding this to the other values in the fours column results in a 3 in the fours column (because the two numbers being added together each have a 1 in the fours column), so the eights column now has a 1 in it, and a 1 is still left in the fours column. Now listing the columns together reveals the answer in the binary form: 1101. Finally, the calculator translates this answer back into decimal form and displays it on the screen: $8 + 4 + 0 + 1 = 13$. As illustrated

by this example, the columns in the binary numbering system cause each other to increase much quicker than the columns in the decimal number system. Many calculators use this form of addition as the basis for the most complicated of operations.

Most calculators allow combinations of operations, but paying attention to the order of operations is essential. For example, a calculator can find the value of four plus six and then divide by two to arrive at five, or it can find the value of four plus the value of six divided by two to arrive at seven. If the numerical and operational numbers (e.g., addition and division) are pressed in the wrong order, the (correct) answer to the wrong question will be found. For example, adding two numbers, dividing by two, and then adding another number usually results in a different value than adding three numbers and then dividing by two.

The ability to store numbers is a valuable function of a calculator. For example, if it takes a long series of operations to find a number that will be used in future calculations, the number can be stored in the calculator (usually by pressing a button labeled STO) and then recalled when needed (usually by pressing a button labeled RCL). Some universally important numbers have been permanently stored in most calculators. Most scientific calculators, for example, have a button for recalling a reasonable approximation of the value of pi, the number that defines how a circle's radius is related to its circumference and area. The exact value of pi cannot be represented on a typical scientific calculator, and repeatedly typing in the numbers involved in the approximation of pi would be tedious to say the least. The ability to quickly provide important numbers is one of most significant benefits of electronic calculators.

Calculators that are capable of more than basic addition, subtraction, multiplication, and division usually have the ability to work in three different modes: degrees, radians, and gradians. These modes pertain to different units for measuring angles. Degrees are used for most of the basic operations. A right angle is 90 degrees and a circle encompasses 360 degrees. Radians measure angles in terms of pi, where pi represents the same angle as 180 degrees (a straight line or half way around a circle). Most calculators indicate that they are working in the degree mode by displaying DEG in the screen. A right angle is half of pi and a circle is represented by pi multiplied by two. When a calculator is working in terms of radians, RAD usually appears in the screen. In gradians, a circle is represented by 400; so a right angle is 100 gradians and a straight line is 200 gradians. This mode is usually indicated by GRAD displayed in the screen.

A Brief History of Discovery and Development

As previously mentioned, the decimal numbering system is based on the number ten because the earliest calculating devices were the ten fingers found on the human body. As human intelligence developed, calculators evolved to incorporate pebbles and sticks. In fact, the word calculator comes from a form of the late fourteenth century word calculus, which originally referred to stones used for counting. Long before the inception of the word, many different ancient civilizations used piles of stones (as well as twigs and other small plentiful things) to count and perform basic addition. However, counting out large piles of stones had limitations (imagine counting 343 stones and then adding 421 stones to find the sum). As civilizations progressed, needs for more efficient calculators increased. For example, more and more merchants were selling their goods in the growing towns, and keeping track of sales transactions became a common need.

Around 300 B.C., the Babylonians used the first counting board, called the Salamis Tablet, which consisted of a marble tablet with parallel lines carved into it. Stones were set on each line to indicate how many of each multiple of five were needed to represent the number. Counting boards similar to the Salamis Tablet eventually appeared in the outdoor markets of many different civilizations. These counting boards were usually made of large slabs of stone and intended to remain stationary, but people with more money could afford more portable boards made of wood.

The abacus took the counting board methods to another level by allowing beads to be slid up and down small rods held together by a frame. The word abacus stems from the Greek word abax, meaning table, which was a common name for the counting boards that became obsolete with the popularization of the abacus. Historians believe that the first abacus was invented by the Aztecs between A.D. 900 and 1000. The Chinese version of the abacus, which is still the calculator of choice in many parts of Asia, first appeared around A.D. 1200. In A.D. 1600, a Russian form of abacus was invented. A Japanese style of Abacus was invented in 1930 and is still widely used in that country. The rods of most abaci are divided into two sections (called decks) by a bar, with the beads above the bar representing multiples of five. A top bead in the ones column represents five, a top bead in the tens column represents 50, and so on. Some abaci have more than two decks. In 1958, the Lee abacus was invented by Lee-Kai-chen. This abacus is still used in some areas. It can be thought of as two abaci (the plural

View of the inside of the first miniature calculator, invented at Texas Instruments in 1967. CORBIS/SYGMA.

of abacus) stacked on top of each other, and is supposed to facilitate multiplication, division, and other more complicated operations.

Mathematical tables and slide rules were two of the most common computational aids before small electronic calculators became reasonably affordable in the 1970s. Mathematical tables were used for thousands of years as a convenient way to find values of certain types of mathematical problems. For example, finding the value of 23 multiplied by 78 on a multiplication table only requires finding the row next to the number 23 and then following that row until reaching the column labeled 78; no computation is necessary, and finding the value takes little time.

The first slide rule was created in 1622. A typical slide rule consists of a two or more rulers marked with numeric scales. At least one of the rulers slides so that two or more of the scales move along each other. Different types of slide rules can be used to reduce various complex operations to simple addition and subtraction. By aligning the scales in the proper positions and observing the positions of other marks on the rulers, a trained user can make quick computations by reducing multiplication and more complex operations to simple addition. Slide rules, along with mathematical tables, remained two of the most useful mathematical tools until they were made obsolete in most areas of computation by the invention of electronic calculators.

The invention of the slide rule was dependent on the discovery of logarithms about a decade earlier because the scales on a slide rule involve logarithms. John Napier was the first to publish writings describing the concept of logarithms, though historians also point out that the idea was most likely conceived a few years earlier by Joost Bürgi, a Swiss clockmaker. The math behind the discovery and development of logarithms is beyond the scope of this text, but their main contribution to science and mathematics lies in their ability to reduce multiplication to addition, division to subtraction. Furthermore, exponents can be found using only multiplication; and finding roots only involves division. For example, when using a table of logarithmic values to multiply two large numbers, one only needs to find the logarithmic values for both of the numbers and add them together. The invention of the slide rule made it possible to work with logarithms without searching through large tables for values.

Many mechanical calculators were invented before the electronic technology used in modern calculators came about. One such mechanical calculator, the Pascaline, was invented in 1642 by 19-year-old French mathematician Blaise Pascal. The Pascaline was based on a gear with only one tooth attached to another gear that had ten teeth. Every time the gear with one tooth completed a turn it would cause the other gear to move a tenth of the way around, so the gear with ten teeth completed one turn for every ten turns of the gear with one tooth. Using multiple gears in this way, the Pascaline mechanically counted in way similar to a person counting on their fingers or using an abacus. The concepts first explored in the Pascaline mechanical calculator are still used in things like the odometer that keeps track of how far an automobile has gone, and the water meter that keeps track of how much water is used in a household.

Compact electronic calculators were made readily available in the early 1970s and changed mathematics forever. Not only were these calculators small and easily portable, they substituted for both slide rules and mathematical tables with their ability to store important and commonly used numbers and to use them in complex operations. With clearly labeled buttons and a screen that shows the answer, these calculators were easier to use and required less practice to master. Like slide rules, many modern electronic calculators use logarithms to reduce mathematical operations to repeated operations of addition.

Personal computers are powered by the same type of technology as handheld calculators. Most computers include a software program that simulates the look and

feel of a handheld calculator, with buttons that can be clicked with the mouse. The main difference between computers and calculators is that computers are capable of handling complex logical expressions involving unknown values. This basically means that computers are capable of processing more types of information and performing a wider variety of tasks. Making the jump from calculators to computers is an important technological milestone. Just as people a thousand years ago could not have imagined a small battery-operated mathematical tool, it is difficult to imagine a technology that will replace electronic calculators and computers.

Real-life Applications

FINANCIAL TRANSACTIONS

When it comes to personal finances, electronic calculating devices have gone far beyond helping people balance checkbooks. Cash registers and automatic teller machines (ATMs) have shaped how people trade money for products and services.

Cash Registers A cash register can be thought of as a large calculator with a secured drawer that holds money. The cash register was originally invented in 1879 to prevent employee theft. The drawer on most cash registers can only be opened after a sales transaction has taken place so that employees can not purposely fail to record a transaction and pocket the money. Manually opening the drawer requires either a secret code or a key that is kept safe by the store manager or owner. The buttons on a cash register are different from the buttons on calculators intended for personal use. The basic buttons of a calculator that are applicable to money (e.g., the numbers and the decimal point) are present on a cash register; but the remaining buttons can usually be customized to fit the needs of the organization that uses it. For example, a restaurant can program a group of buttons to store the prices of their various menu items; or cash registers in certain geographic locations might have buttons for computing the regional sales tax. The screen can usually be turned so that the merchant and the customer can both see the prices, taxes (if any), and total. Like many calculators, a cash register has a roll of paper and a printing device used for creating printed records of calculations (called receipts in the case of monetary transactions). The inside of a cash register works (and always has worked) almost exactly like a calculator. Modern cash registers include electronic microprocessors similar to those found in handheld calculators; but when calculators were powered by the turning of mechanical gears, cash registers were also powered by similar gear mechanisms.

ATM Machines Automatic teller machines (ATMs) were first used in 1960 when a few machines were placed in bank lobbies to allow customers to quickly pay bills without talking to a bank teller. Later in the decade, the first cash dispensing ATMs were introduced, followed by ATMs that could accept and read bank cards. The fact that ATMs are unmanned requires that they possess greater security. To ensure the safety of the bank's money, the materials that make up the ATM and connect it to a building are precisely constructed and physically strong. To thwart attempts to pose as another person in order to take that person's money out of an ATM, transactions require two forms of identification: physical possession of a bank card and knowledge of a personal identification number (PIN). While the inner workings of an ATM are more complicated than that of a cash register, the technology and concepts of the electronic calculator provide the basis for computing the values of every transaction.

The introduction of check cards has combined the technological benefits of cash registers and ATMs to further facilitate the storage and expenditure of money. A check card can be used to make purchases using money that is stored in a checking account at a bank in another location. Other advancements in technology (e.g., scanners that quickly scan barcodes on items, self-checkout stations that allow customers to scan their own items, and secure Internet transactions that use calculators operating on a computer thousands of miles away from the computer being used by the customer) continue to revolutionize how humans buy and sell products and services. However, none of these accomplishments would have been possible without tools that automatically perform the mathematical operations that take place in every monetary transaction.

NAUTICAL NAVIGATION

For hundreds of years, sailors used celestial navigation: navigating sea vessels by keeping track of the relative positions of stars in the sky. Through the ages, a wide variety of tools have been created to help a navigators navigate boats and ships from one point to another in a safe and timely manner. Different colored buoys warn of shallow waters or fishing nets, and ensure that ships do not collide when nearing docks and harbors. A compass is an essential tool for determining and maintaining directional bearings. Tables of tides and detailed nautical maps help to determine the quickest and safest route and foresee potential obstacles and dangers. For centuries,

Beat the Abacus

Contests throughout the world have pitted individuals equipped with an abacus against individuals equipped with a handheld digital calculator. In most cases, the person with the abacus wins, no matter how complicated the mathematical operations involved. This, of course, does not mean that even the most skilled person with an abacus can make calculations faster than a calculator; the time that it takes to press the buttons accounts for most of the time that it takes to use a calculator to solve a problem. Nonetheless, even in operations as complicated as multiplying and dividing 100 pairs of numbers with up to 12 digits (trillions), a proficient abacus user beats a skilled calculator user almost every time.

navigation of the seas required an in-depth understanding of trigonometry (relationships between lengths and angles) and intensive calculations performed by hand; and, as many navigators have discovered the hard way, small directional errors can result in devastating miscalculations over a trip of thousands of miles. Handheld electronic calculators have proven to be an essential navigational aid since they became reasonably affordable. They are often used aboard sea vessels as either the primary tool for calculating directions and distances on the water or the secondary tool for double-checking calculations carried out by hand.

For every type of navigational problem that can be solved with the help of a handheld electronic calculator, there is also a specialized calculator for solving the specific problem. Often found either on a sea vessel or on the Internet, several calculators have been programmed to take a few pertinent values and find a specific answer. One example of a specialized nautical calculator is a speed-distance-time calculator for finding the time that it will take to get from one point to another if traveling at a certain speed. Most of these calculators require two of the three values (speed, distance, and time) in order to calculate the third value. The time that it takes to get from one point to another is the product of the distance between the two points and the speed at which the ship is traveling (time is equal to distance multiplied by speed). Similarly, to figure out how fast the ship needs to travel in order to get from one point to another in a specified amount of time requires dividing the distance by the desired time of travel (speed is equal to distance divided by time). Finally, to figure out far a ship will go if traveling at a given speed for a specified amount of time, the speed and time must be multiplied together (distance is equal to speed multiplied by time). Due to the fact that all of these operations involve only multiplication and division, this type of calculator only needs to be capable of multiplication and division. More sophisticated navigation calculators exist to quickly determine values that help a ship's navigator make crucial decisions. These decisions range from determining the fuel necessary for completing a trip and planning appropriate stops for refueling, and finding the true direction in which to steer the ship in order to maintain a desired heading (direction) while taking into account forces such as wind and the current of the water. Specialized calculators are also often used to ensure that a ship is built properly. One such calculator measures a ship's resistance to capsizing (turning upside-down in the water) based on the width of the widest part of the ship and the weight of the ship.

Although modern global positioning system (GPS) technology allows precise and accurate position measurements, calculators (whether external or internal) are used determine vectors (directions and distance) to execute course changes or to determine the best path.

COMPOUND INTEREST

Banking can be a highly profitable business. For example, a bank can use the money in a savings account for other investments as long as the money is stored at the bank; so the more money present in the bank's various accounts at any given time, the more money the bank can earn on its own investments. As an incentive for banking customers to store their money with a bank, savings accounts earn compound interest. That is, the bank pays a savings account holder a relatively small amount of money based on the amount of money in the savings account. The basic idea that drives this investment chain is that the bank makes more money in its own investments than it pays out to its account holders.

The amount of money that is earned on a savings account containing a given amount of money is determined by a compound interest formula. Compound interest is an example of exponential growth: the larger the number becomes, the faster the number grows. The term compound refers to the idea that the growth depends both on how much money is deposited into the account as well as the amount of interest already earned in past growth periods. These growth periods are referred

to as compounding periods. Interest is typically compounded annually or monthly, but may also be compounded weekly, or even daily. More frequent compounding benefits the account holder and may be offered to attract more account holders in order to increase the bank's profits.

Determining the amount of interest earned and predicting future account values requires calculations of inverses (1 divided by a number) and exponents (one number raised to the power of another number), both of which are usually rather messy operations, especially when performed by hand. A handheld scientific calculator allows account holders to calculate these values quickly and accurately in order to compare banks and track earnings with ease.

MEASUREMENT CALCULATIONS

How calculators function to solve an array of measurement and conversion problems is perhaps best illustrated by example. Imagine a local high school is hosting a regional basketball tournament. On the day of the tournament, the athletic director discovers that the supply closet has been vandalized and all of the basketballs have been damaged. As the athletic director begins to make the announcement that the tournament will have to be delayed due to the lack of basketballs in the building, a student in the crowd reveals that she has a basketball in her backpack and throws it down to the court. Before the tournament can resume, the officials must determine whether or not the ball is regulation size. All of the writing, including the size of the ball, has been worn off by years of use. Fortunately, one of the referees knows that the diameter of a full-sized basketball (the distance from one side of the ball to the other measured through the center of the ball) is about 9.4 inches. The high school home economics teacher, who happens to be in the crowd, quickly produces a tape measure from her purse, hoping to be of assistance. However, an accurate measurement of the diameter of the basketball cannot be determined with a tape measure. The referee measures the circumference of the ball (the longest distance around the surface of the ball) and finds that it is 29.5 inches. Not knowing the circumference of a regulation-size basketball, the referee asks if anyone in the crowd might know how to solve this problem.

A student speaks up, stating that he has been studying circles and spheres in his math class. He was able to recall an important fact that would help to determine the diameter of the basketball: the circumference of a sphere (such as a basketball) is equal to the diameter of the sphere multiplied by pi. So the diameter of the basketball is 29.5 divided by pi. The student cannot remember a good approximation of the value of pi, but his scientific calculator has a button for recalling the value of pi (approximated to the ten digits that his calculator can display). He enters 29.5, presses the / button (for division), recalls the value of pi (which displays 3.141592654), and presses the = button (the equal sign). The calculator displays the answer as 9.390141642. This value rounds to 9.4, which is the value that the referee indicated as the diameter of a regulation basketball. The ball is accepted by the officials and the tournament continues.

RANDOM NUMBER GENERATOR

When conducting scientific experiments, it is often necessary to generate a random number (or a set of multiple random numbers) in order to simulate real-life situations. For example, a group of scientists attempting to model the way that fire spreads in a forest need to account for the fact that a burning tree may or may not ignite a nearby tree. Unpredictable factors like shifting winds and seasonal levels of moisture make incorporating the probability of fire spreading in a certain direction into models next to impossible because the nature of wildfires is seemingly random. However, this randomness can be loosely accounted for in scientific wildfire models by strategically inserting random numbers into the mathematical formulas that are used to describe the nature of the fire. These models are often run repeatedly in order to evaluate how well they fit real-world observations. Each time the formula is used, different random numbers are generated and inserted into the formula.

Cryptography Another important area of study that benefits from the generation of random numbers is cryptography, in which messages are encrypted (scrambled) so that they cannot be understood if they are intercepted by an unauthorized party. A message is encrypted according to mathematical formulas. Most of these encryption formulas incorporate random numbers in order to create keys that must be used to decrypt (unscramble) the message. The decryption key is available only to the message sender and the intended message reader.

Random number generators are important tools in many other scientific endeavors, from population modeling to sports predictions. Fortunately, most scientific calculators and graphing calculators include buttons for generating random numbers. Some calculators have a single button (often labeled RAND or RND) for generating a random three-digit number, between 000 and 999. Each time the button is pressed, a new random number is generated. Other calculators also allow the user to adjust the number of digits and the placement of the decimal

point in the random generated numbers. Other calculators allow the user to define upper and lower bounds for random numbers; and some can generate multiple random numbers at once. On such a calculator, inputting a set of three numbers that looks something like (1, 52, 9) will cause the calculator to display nine random numbers with values between one and 52. These values can then be used to represent the drawing of nine cards from a deck of 52 playing cards, where each card is assigned a number between one and 52.

Random number generators included in calculators (and various computer software programs) not only make it easy to generate the random numbers needed to simulate real-life situations; using random number generators also ensures that the numbers are truly random. The idea of structured randomness may seem strange; but in order to fully simulate the true randomness found in real-life situations, random number generators use mathematical formulas to generate the numbers according to certain guidelines. One such guideline ensures that the numbers are distributed in certain ways (e.g., to ensure that the numbers are not all close together or equally spaced). Different methods for achieving randomness are used to generate random numbers, and choosing a method is an important consideration in scientific modeling scenarios. Random numbers generated according to mathematical formulas are referred to as pseudorandom numbers.

With the hordes of numbers and unknowns required to model real-life situations, it can be easy to lose sight of the essential ideas behind the data. Using random numbers in mathematical models makes it possible to imitate experiments and focus on the underlying patterns and ideas.

BRIDGE CONSTRUCTION

The construction of large suspension bridges requires almost unfathomable amounts of calculations to ensure that the structures can withstand the multitude of forces acting on a bridge at any given time. Although a suspension bridge looks solid, it is a complex structure that is constantly swaying and twisting; if it were rigid, it would snap under heavy winds and other forces. The weight of the roadway alone would cause the bridge to crumble if swooping cables attached to strong towers were not accurately designed and built. Some forces, including gravity and the weight of the materials that make up the bridge, are constant (unchanging). Other factors are constantly changing: the weight of the automobiles, the strength of the wind, the strength of the water current pushing on the supporting structure, varying temperatures, earthquakes and other disastrous activity. Bridge engineers must ensure that a bridge can withstand the worst possible situations. For example, a worst-case calculation might examine the stability of the bridge supporting the maximum number of automobiles while under the pressure of high winds and strong water currents during a reasonably large earthquake. When building a large bridge, the slightest miscalculation has the potential of endangering hundreds of human lives.

Before the invention of electronic calculators, the colossal calculations involved in building a safe and long-lasting bridge were performed (and rechecked many times) by hand with the assistance of slide rules and enormous mathematical tables. When the Golden Gate Bridge was built in San Francisco, California, it was the longest suspension bridge in the world. Most experts believed the distance that needed to be spanned in order to build a bridge across the Golden Gate Straight was too large. Furthermore, the many other regional complications—including characteristically high winds, strong tidal currents, the weight of water formed by dense fog, and frequent earthquake activity—made most bridge engineers skeptical to say the least. However, Joseph B. Strauss, who worked on hundreds of bridges in his life, successfully planned and headed the construction of the Golden Gate Bridge. Strauss and his team of engineers worked for months using circular slide rules and making (and rechecking) calculations involving more than 30 unknowns (e.g., the height of the towers, the lengths and arcs of the cables, the thickness of the roadway, the speed of the wind, the strength of water currents, and the weight of automobiles). The bridge took over four years to build, spanned 4,200 feet (1,280 m), and cost over 30 million dollars. To someone accustomed to using a handheld electronic calculator, even the task of approximating the cost of the bridge—taking into account the amounts of materials, the number of people required for construction, and the predicted amount of time needed—seems daunting.

The invention of electronic calculating devices greatly reduced the amount of time needed to perform and repeat immense calculations. For example, the stability of a suspension bridge depends heavily on the lengths of the cables, the heights of the towers to which the cables are connected, and the angles between them. A typical scientific calculator has buttons labeled SIN, COS, and TAN. These buttons are related to trigonometry, the study of triangles that defines the relationship between lengths and angles, and greatly reduce the time needed to calculate and confirm crucial measurements for the parts of a bridge.

Bridge engineering continued to advance as calculating devices evolved into computing technology that could

quickly and accurately simulate diverse situations involving numerous adjustable factors. The record for longest suspension bridge has been broken many times since the completion of the Golden Gate Bridge. In 1998, the Akashi Kaikyo Bridge was constructed across the Akashi Strait between Kobe and Awaji-shima in Japan. This massive steel bridge was the longest (and tallest) suspension bridge as of 2005, spanning a total of 12,828 feet (3,910 m). It took over ten years and 4.3 billion dollars to build.

COMBINATORICS

Combinatorics is the mathematical field relating to the possible combinations of a given number of items. A common example investigates the number of possible arrangements (called permutations) for a standard deck of 52 playing cards. It can be shown that 52 cards can be arranged in a surprisingly large number of ways, which is essential in making cards games unpredictable and interesting.

To grasp the idea, start with the order in which cards are usually organized when the pack is first opened: increasing from two to king (with the ace on one end or the other) and separated by suit (hearts, spades, diamonds, clubs, not necessarily in that order). That's one combination (permutation). Next, take the card on the top of the deck and move it down one position. That's two combinations. Continue to move that card down one position until it reaches the bottom of the deck. That is 52 different combinations obtained by moving a single card. Next, take the next card in the deck (the card that is now on the top of the stack) and move it through all of the possible positions. Keep in mind that the first combination was already accounted for when the first card was in its final position on the bottom of the stack; so that is another 51 combinations of cards. The next card will provide another 50 combinations, and so on. It turns out that the total number of combinations is the product of the numbers between 52 and one: 52 multiplied by 51 multiplied by 50, and so on down to one. This type of value (the product of every whole number between a given number and one) appears often in combinatorics and has a standard notation. The number of combinations for 52 cards, for example, is written as 52! and pronounced fifty-two factorial. This type of multiplication is difficult and time-consuming to work out by hand. Fortunately, typical scientific calculators include a button for performing factorial operations (usually labeled n! and pronounced n factorial). Entering 52 and then pressing n! returns a number larger than eight multiplied by ten to the 67th power. That means that number of possible combinations for a deck of 52 cards is more than 8 followed by 67 zeros! A million has only six zeros; a billion only nine. Factorial operations tend to yield large numbers and are difficult to calculate by hand; but calculators make it easy to find the values of factorials of reasonably large numbers, and even perform operations on those values.

UNDERSTANDING WEATHER

The practice of predicting the weather has been a growing art for centuries; but no advancement has influenced the field meteorology (the scientific study of Earth's climate and weather) more than the development of calculating and computing devices. As with most scientific fields, the common availability of electronic calculators affected meteorological studies by greatly reducing the amount of time required for making calculations needed to predict the weather, and updating these calculations based on frequent changes in observed weather data.

American meterologist Joanne Simpson, the first woman to earn a doctorate in meteorology, developed the first model of cloud activity in Earth's atmosphere, helped to explain the forces that power hurricanes, and discovered the cause of the air currents in tropic regions. The calculations that led to her theories were originally performed in the 1940s and 50s without the assistance of an electronic calculator. Simpson's theories were met with criticism and disbelief, but she would later stand as a shining example of electronic calculating devices verifying human calculations. Using calculators and, eventually, computers, she was able to improve her models, revolutionizing meteorological research and prediction. After years of tireless research and teaching positions at multiple universities, she went on to work at the National Aeronautics and Space Administration (NASA) for over 24 years.

While working at NASA, Simpson was integral in the advancement of meteorological studies using images and information gathered by satellites orbiting Earth. Starting in 1986, Simpson headed NASA's Tropical Rainfall Measuring Mission (TRMM). This mission involved the launch and utilization of the first satellite capable of measuring the rainfall in Earth's tropical and subtropical regions from space. This mission has been regarded as one of the most important advancements in the field of meteorological research, deepening the understanding of meteorological phenomena ranging from the affects of dust and smoke on rain clouds to the origins of hurricanes.

The scientific accomplishments of Simpson—from hand calculations leading to ground-breaking theories, to cutting-edge technological research—provide an excellent illustration of the power of a brilliant mind teamed up with technology. Having already revolutionized her field long before the availability of electronic calculating

Key Terms

Calculator: A tool for performing mathematical operations on numbers.

Decimal: Based on the number ten; proceeding by tens.

Digital: Of or relating to data in the form of numerical digits.

Exponent: Also referred to as a power, a symbol written above and to the right of a quantity to indicate how many times the quantity is multiplied by itself.

Logarithm: The power to which a base number, usually 10, has to be raised to in order to produce a specific number.

Model: A system of theoretical ideas, information, and inferences presented as a mathematical description of an entity or characteristic.

Operation: A method of combining the members of a set so the result is also a member of the set. Addition, subtraction, multiplication, and division of real numbers are everyday examples of mathematical operations.

devices, she first used calculators to verify her results, and then continued to stretch the limitations of meteorology by employing the ever-growing capabilities of computers.

Potential Applications

SUPERCOMPUTERS

The earliest and simplest personal computers differed from calculators in that they could interpret instructions involving unknown values, called variables. As computers evolve, they retain the ability to perform the mathematical calculations for which calculators exist. Most personal computers have a calculator program. On the computer screen, the calculator program resembles a handheld calculator. The buttons on the screen can be clicked with the mouse, or the computer keyboard can be used to input the numbers and commands. Even the most advanced computers are based on the concepts that enable calculators to perform mathematical operations.

Supercomputers are computing systems that possess the most power and are capable of the highest level of computation in any given time period. Since the early days of digital computation, supercomputers have been employed to perform large amounts of complex calculations that go beyond the capabilities of the common computing machines of their era. Supercomputers (also known as high performance computing systems) become outdated as technology advances. For example, the most powerful personal computers in the new millennium possess power more than supercomputers from past technological eras.

Long before handheld electronic calculators were made available to the general public, electronic supercomputers were used by the United States government to break enemy codes during World War II. Military codes are usually implemented using mathematical formulas that define how information is transformed, and in turn, how to transform it back to its original, readable form. Cracking these codes requires massive amounts of calculations to determine the values of the mathematical formulas.

In the 1990s, large oil companies began deploying supercomputers developed to analyze enormous amounts of seismic data (information about vibrations in the ground). This data can be used to create images that represent the underlying contents of the terrain, helping the oil companies locate oil in the ground and leading to a dramatic increase in the accuracy of their searches.

Other uses for supercomputers include the creation of detailed three-dimensional (3-D) models of information that is difficult to grasp in its raw form. Weather patterns can be visualized, making them easier to interpret, hopefully improving future predictions. The chemical compositions of a virus can be seen in a new light that may help medical experts expose a plan of attack. The surface of planets can be recreated based on information gathered by a spacecraft. Supercomputers make it possible to understand the invisible building blocks of our world and to access the far reaches of the universe for close inspection. Like all major scientific contributions, supercomputers allow the men and women of science to achieve goals that were previously nothing more than dreams. Considering the abundance of new applications leading to breakthrough scientific discoveries each year, supercomputers seem to possess boundless potential for enhancing our understanding of science.

Where to Learn More

Books

Tao, W., and R. Adler *Cloud Systems, Hurricanes, and the Tropical Rainfall Measuring Mission (TRMM)-A Tribute to Joanne Simpson*. Boston: American Meteorological Society, 2003.

Periodicals

Petroski, H. "Keeping Swaying Bridges from Falling Down." *Science* no. 295 (2002): 2374–2375.

Web sites

Boats.com. "Navigating With a Calculator: Solving navigation problems with the touch of a button." <http://www.boats.com/boat-articles/Seamanship-148/Navigating+With+a+Calculator/2932.html> (March 18, 2005).

University of Mannheim. "Top 500 Supercomputer Sites." <http://www.top500.org/> (March 16, 2005).

Calculus

Overview

Calculus is a set of mathematical tools for solving certain problems. It assumes the methods of algebra and geometry, and in turn is used as a starting point for higher forms of mathematics. Calculus is applied to a wide range of problems in medicine, the social sciences, economics, biological growth and decay, physics, and engineering. It is a flexible language for describing the physical world: objects in motion, chemical reactions, complex surfaces and volumes, heating and cooling, the hard-to-imagine behaviors of space and time, and many other events. Not only do all students of medicine, engineering, physics, biology, and economics study calculus in college, so do many students of psychology, literature, art, business, and history.

Calculus, like geometry, is built on a foundation of simple elements that can be drawn on paper or seen in the mind: curves, slopes, and areas. And if ever a branch of mathematics was created to deal with real-life problems, this is it. English physicist Isaac Newton (1642–1727) and German mathematician Gottfried Wilhelm von Leibniz (1646–1716) invented calculus in the 1600s because it was needed to solve the cutting-edge science and math problems of their time, including how to calculate the lengths of curves, the areas bounded by curves, and the motion of objects that are accelerating (gaining speed).

More than three hundred years later, the language of calculus is common to almost every scientific or technical field. It is essential to the design of engines, computers, and all other complex machines; to economics, agriculture, and physics; in fact, to the study of pretty much everything from subatomic particles to the shape of the Universe. Almost every scientist and engineer in the world understands at least basic calculus. And even for those who will never as adults perform any mathematical operation more complex than bidding on eBay, the *concepts* of calculus, like the concepts of basic physics, can clarify our thinking about the world, make it more effective.

Geometry tells us how to deal with curves and shapes that are described by simple rules: squares, triangles, circles, ellipses, and so on. Calculus tells us how to deal with curves and shapes that are described by more complex rules, or by no rules. Calculus is more suited to the real world, where irregular curves and shapes—including graphs of changing speeds, forces, and other things that can be measured—are the rule.

Fundamental Mathematical Concepts and Terms

FUNCTIONS

To talk about calculus, we need to be able to talk about functions. A "function" is a rule that relates one group of numbers to another. For example, we can write a rule that relates each positive number, x, to some negative number, $f = -x$. This rule or function tells us that if x equals 10, f equals -10. Likewise, if x equals 2.5 then f equals -2.5, and so on. We say that "f is a function of x" when f is on the left-hand side of the equals sign, as it is here.

When f is a function of x, it is common practice to write f as $f(x)$, which we read aloud as "f of x." So the function $f = -x$ can also be written as $f(x) = -x$. If we do not want to write the whole rule out every time (or if we do not know what that rule is), we can simply write $f(x)$.

By the way, other letters besides x and f can be used to write functions. In fact, if we are talking about more than one function at a time, we have to use other letters to keep from getting confused.

It is often useful to make a picture of a function. This is done by picking values for x, applying the rule of the function, and finding out what values of f result. In this way any number x can be paired with a number f. These pairs can be graphed as dots by hand or computer. If enough of these dots are graphed, they can be joined by a smooth line. In this discussion, such graphs will be used to show what various functions look like. For example, a graph of the function $f(x) = 2x$ is shown in Figure 1.

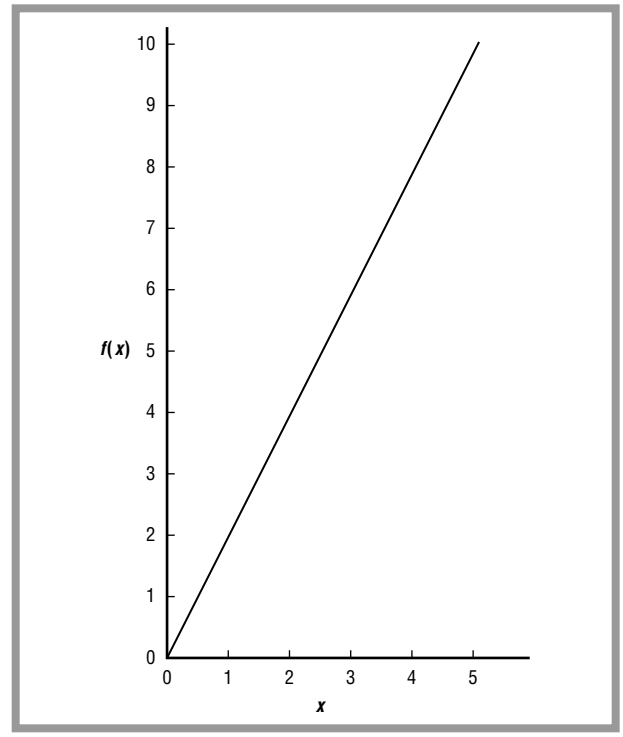


Figure 1: Graph of the function $f(x) = 2x$. The line could go on forever, but only a part can be shown.

THE DERIVATIVE

Consider the simple function shown in Figure 2, $f(x)$. The exact rule that relates f to x is not important right now; what matters to us is the shape of its graph.

The function depicted is often used in real-life math problems. It approximates one of the curves used to describe the spread of epidemics and other processes that spread geometrically in a finite medium (discussed in the article on Exponents). It also approximates part of the spectrum of tidal wetland growth (discussed further below as a real-life application of derivatives). Such a curve is shown in many introductory calculus textbooks because it offers a clear visual basis for explaining the concepts of calculus. In particular, tangent lines with positive slope can be laid against the curve without confusingly overlapping over other parts of it; also, the integral of (area under) this curve is easy to understand and to graph. Moreover, the derivative of this curve is a nontrivial bell shape that can be used to explore min-max problems.

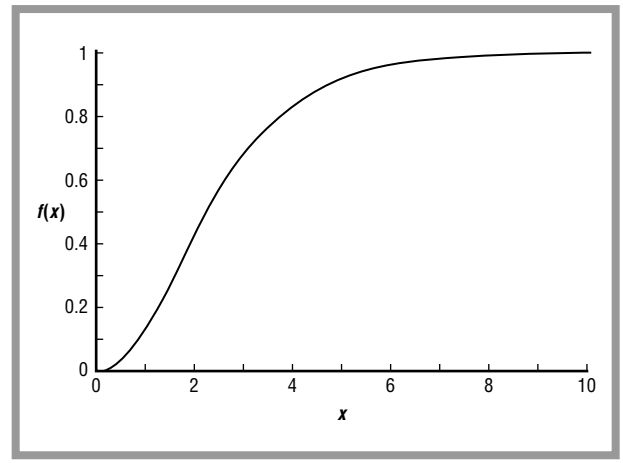


Figure 2: A simple function.

We will use this function to visualize the first ideas of calculus, but the questions we are about to ask about this particular $f(x)$ could just as well be asked about many other curves.

The two basic ideas of calculus arise from asking two questions about curves like this one. The first is: How steep is the curve at any one point? A more exact way of

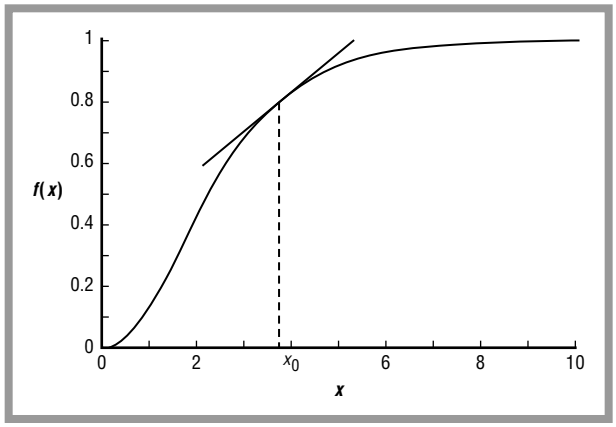


Figure 3: A line segment tangent to $f(x)$.

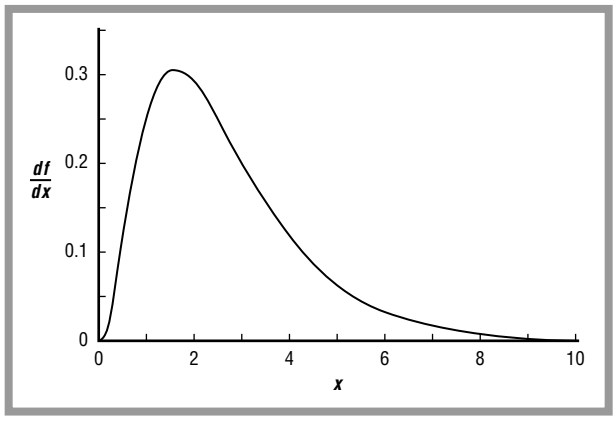


Figure 4: Derivative of the function in Figure 2.

asking this is: If we lay a straight line against the curve so that it touches it at just one point, how steep is that line? In other words, what is its slope? Figure 3 shows a line segment just touching $f(x)$ at the point directly above x_o. (The subscript "0" is just a label to distinguish x_o from other values of x.) This line segment is said to be "tangent" to the curve.

The slope of a line tangent to a curve is called the *derivative* of the curve at the point where the line and the curve touch. The derivative tells us how steep the curve is where the line touches.

A single tangent line shows the derivative only at one point, but the derivative can be found in the same way for every single point along the curve. These numbers can be graphed as a curve in their own right. From Figure 3 you can see that the slope of $f(x)$ starts out small (at zero, actually), gets bigger as $f(x)$ increases, and shrinks as $f(x)$ starts to level off. The derivative of $f(x)$ is shown in Figure 4.

The derivative of $f(x)$ is often written

$$\frac{df}{dx}$$

because it corresponds to the slope or rate of change of $f(x)$ at a single point x. The numerator df stands for a very small vertical change in $f(x)$, and the denominator dx stands for a very small horizontal change in x, so

$$\frac{df}{dx}$$

echoes the definition of a slope from elementary geometry,

$$\frac{\text{rise}}{\text{run}}.$$

Another way of looking at the derivative or slope of a function at a given point is that it tells you the *rate of change* of the function at that point.

A function $f(x)$ might be defined either by a series of measurements of some real-world quantity, or by an equation. In the case of the curve in Figures 2 and 3, the equation for $f(x)$ happens to be $f(x) = (1 - 2^{-x})^3$. There is a set of standard rules (which students in introductory calculus courses learn by heart) that says exactly how to write down what df/dx is, starting with an equation for $f(x)$. Applying these rules to $f(x)$ is called "differentiating" $f(x)$ or "taking the derivative of" $f(x)$. Some functions do not cooperate with these rules and so their derivatives cannot be written down explicitly, in which case computers must be used to find their derivatives.

Taking derivatives is one of the two fundamental operations of calculus. But what use are derivatives? Why bother with them?

Derivatives can help pilot remote vehicles (e.g., to help robot rovers navigate on other planets, such as Mars). If, for example, an engineer is piloting a robot rover on Mars by remote control and plans to drive it up a hill, he might rely on orbital photography to provide data for a graph of altitude versus distance along the rover's proposed line of travel. This graph might look something like the function in Figure 2. But if the Mars rover isn't strong enough to climb at any angle steeper than, say, 30 degrees, the pilot would want to look at the *derivative* of the altitude curve to make sure that the steepness of the proposed route never exceeds 30 degrees at any point. The peak value of the curve in Figure 4 would tell you exactly what maximum steepness your rover was going to encounter. If that value was too high, you'd have to try another route.

One more word about derivatives. The derivative of a function is just another function, and so you can take its

derivative too. This function is called the "second derivative of $f(x)$" and is written

$$\frac{d^2 f}{dx^2}$$

The second derivative of the $f(x)$ graphed in Figure 2 is shown in Figure 5.

If you take the derivative of the second derivative, you have the third derivative of $f(x)$. You could find the fourth, fifth, or ten-thousandth derivative of $f(x)$, too, but the first and second derivatives are by far the most useful. For instance, the first derivative of a function that describes an object's position describes the object's speed, and the second derivative describes the object's acceleration. Any derivative beyond the first is called a "higher-order" derivative.

THE INTEGRAL

Now to ask a simple but important question about the function in Figure 2: What is the area under any given part of the curve? In Figure 6, the area under the curve between $x = 0$ and $x = x_0$ has been shaded in.

This area is called the definite integral of $f(x)$ between 0 and x_0. The definite integral, like the derivative at a single point, is simply a number. In our example, the definite integral says how many square inches the shaded area in Figure 6 is.

The definite integral can tell us the actual physical area of an object with curving edges. It can have other physical meanings, as well. For example, the integral of an equation that describes an object's velocity tells us how far the object has traveled. Consider an object moving at a steady speed or velocity, v. Velocity might change over time, so we will write v as a function of time, $v(t)$. If the object's velocity happens to be an unchanging 100 miles per hour, we can write $v(t) =$ 100 miles per hour. This function is shown in Figure 7.

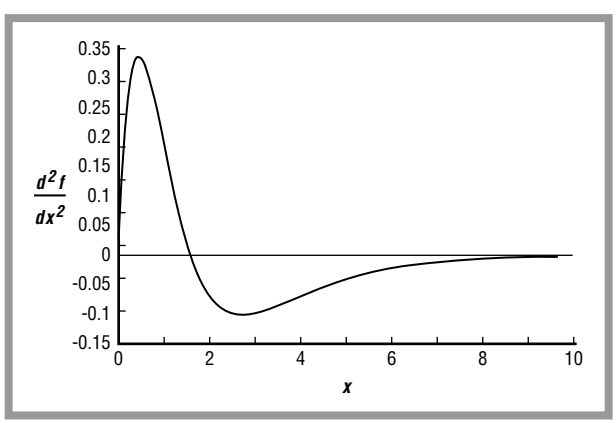


Figure 5: The second derivative of the $f(x)$ shown in Figure 2.

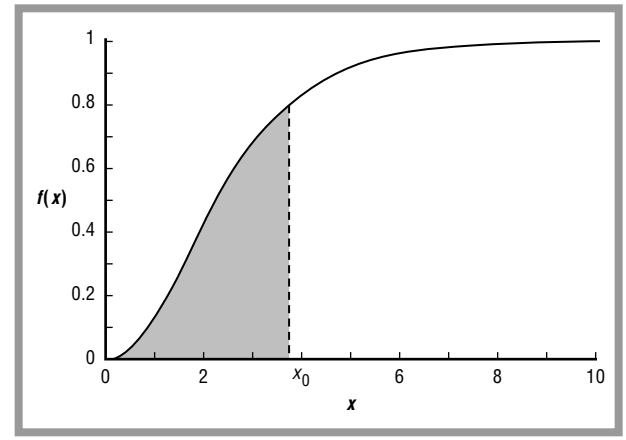


Figure 6: Shaded part is the area under $f(x)$ between 0 and x_0.

The definite integral of the velocity, $v(t)$, from time 0 to time t_0 is the area under the curve from 0 to time t_0. This area is shaded in Figure 8.

The length of the rectangle in Figure 8 is t_0 and its height is 100. Therefore, its area—the value of the definite integral—is $100 \times t_0$. The definite integral of a velocity function is useful because it gives the distance traveled in that time. In 1 hour, the object in our example will have traveled $100 \times 1 = 100$ miles; in 2.5 hours, it will have traveled $100 \times 2.5 = 250$ miles.

Here the area calculation is simple because the velocity is unchanging, so we can use the formula for the area of a rectangle. In real life, objects such as cars, bullets, spacecraft, and runners change their velocities over time. In this case the velocity curve is not a flat line (as in Figure 7) but a more complicated curve, perhaps like that in Figure 6. The more complex the curve, the more complex the mathematics needed to find its integral.

Just as with the derivative, it is possible to find a series of definite integrals and to graph them as a function in their own right. This is called *integrating* $f(x)$, and the resulting curve or function is called the *indefinite integral* (or simply the *integral*) of $f(x)$. Also, instead of graphing the integral point by point by evaluating definite integrals, it is often possible to find an exact expression for the integral. The indefinite integral of a function $f(x)$ is written as follows:

$$\int f(x)\,dx$$

The symbol at the far left that looks like a stretched "s", $\int$, actually *is* a stretched "s." Centuries ago it stood for "summa," which is Latin for "sum," a reference to summing up the area under the curve. This symbol is called the *integral sign*. The expression

$$\int f(x)\,dx$$

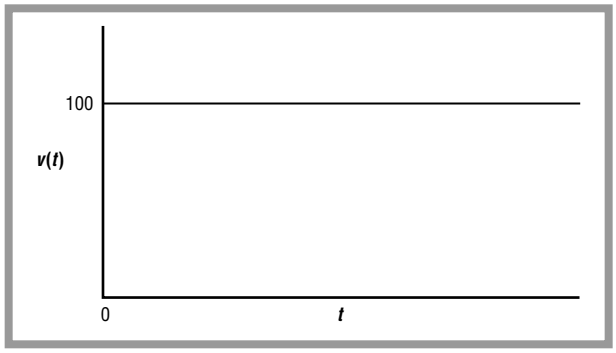


Figure 7: The velocity, $v(t)$, of an object moving at 100 miles per hour.

Figure 8: The area of the shaded rectangle is the definite integral of the velocity $v(t)$ from time 0 to time t_0.

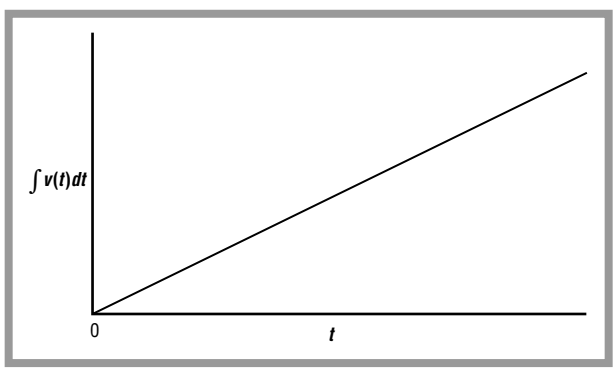


Figure 9: The indefinite integral of $v(t)$.

is read aloud as "the integral of $f(x)$ dee x." (We can use letters other than f and x whenever we like; they are just labels or names.)

In our example of an object moving at a constant 100 miles per hour, the integral of $v(t) = 100$ is easy to write down as an exact mathematical expression:

$$\int v(t)\,dt = 100t + C$$

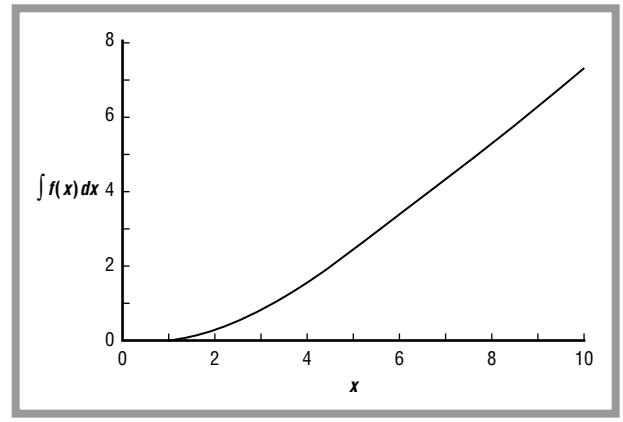


Figure 10: The indefinite integral of the function first seen in Figure 2, $\int kt\,dt = \frac{k}{2}t^2$.

The C at far right stands for "constant of integration." C is "arbitrary," meaning that it can be set equal to any number and the equation will still be true. The indefinite integral of $v(t) = 100$, namely

$$\int v(t)\,dt = 100t + C$$

is plotted in Figure 9 with C set equal to 0.

An indefinite integral is more informative than a definite integral because it can tell us the value of any definite integral. To find the value of a definite integral over a certain interval (for example, from time 0 to time t_0), we subtract the value of the indefinite integral at the left-hand end of the interval from its value at the right-hand end. In the case of the object moving at 100 miles per hour, the value of the indefinite integral at time t_0 is $100 \times t_0 + C$. At time 0 it is $100 \times 0 + C$. Subtracting, we have

$$\begin{array}{r} 100 \times t_0 + C \\ -\,100 \times 0 + C \\ \hline 100 \times t_0 \end{array}$$

which is exactly what we found by calculating this definite integral as the area of a rectangle.

We have already seen, in Figures 4 and 5, the first and second derivatives of the curve first shown in Figure 2. The integral of that curve is shown in Figure 10.

As mentioned earlier, in a real-life application, the "area" under a curve may not correspond to a literal, physical area like 10 square miles of parking lot. In calculus,

depending on what real-world quantity you're measuring, an "area" may be the number of miles driven, or the profit made by a business, or the amount of oil leaked from a beached tanker, or the probability that a rocket will explode before reaching orbit, or many other things. Derivatives are flexible in the same way. It's one of the reasons calculus is so useful.

THE FUNDAMENTAL THEOREM OF CALCULUS

The fundamental theorem of calculus is simply this: Taking the derivative is the reverse of taking the integral, and taking the integral is the reverse of taking the derivative. We have seen this sort of thing before: add 10 to a number and then subtract 10, and you are back where you started (addition and subtraction can reverse each other). Or multiply a number by 10 and then divide it by 10, and you are back where you started (multiplication and division can reverse each other).

Similarly, take the derivative of a function and then take the integral of that derivative, and you are back where you started. (Except for a constant of integration, and that can always be set to whatever we choose.) Integration and differentiation reverse each other.

Writing this out in symbols, we have

$$\frac{d}{dx}\int f(x)\,dx = f(x)$$

(taking the derivative undoes integration) and

$$\int \frac{df}{dx}\,dx = f(x)$$

(integration undoes taking the derivative). (For simplicity, the constants of integration on the right-hand sides of these two integrals, and of the rest of the indefinite integrals in this chapter, are omitted, as is often done in practical math.)

The fundamental theorem of calculus is fundamental because it tells us that a function, the derivative of that function, and the integral of that function all contain the same information in different forms. Knowing any one form, we can produce the others.

And that's it, that's calculus. Or the heart of it, at least: derivatives, integrals, and the fundamental theorem that ties them together. Related topics such as summations, limits, exponential functions, and analytic geometry are often lumped together under the term "calculus" for convenience, but differentiation, integration, and the fundamental theorem are the big three, the core.

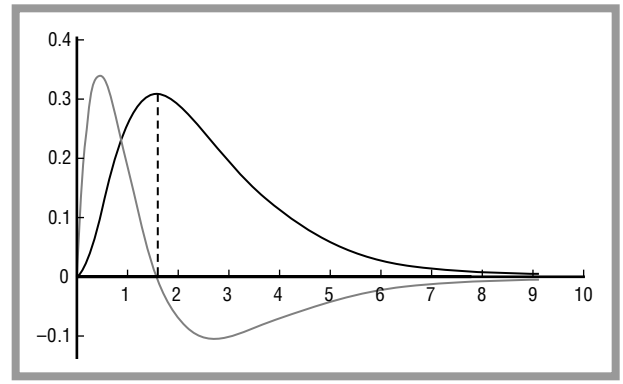


Figure 11: A function (black) and its derivative (gray) plotted together, showing that at the x value where the function levels off at a maximum, its derivative equals 0.

MAXIMA AND MINIMA

The curve in Figure 4 and its derivative, the curve in Figure 5, are plotted together in Figure 11. Notice that the place where the curve from Figure 4 (the black line) crosses the horizontal axis, around $x = 1.6$, the curve from Figure 3 (the gray line) hits a peak or maximum.

This gives us a very useful general principle: By finding out where the derivative of a function equals zero, we can determine the maximum (and minimum) points of that function. Why? Because maxima and minima are peaks and valleys, places where a function levels off briefly. And wherever a curve is level, its slope (derivative) equals zero. Being able to find maxima and minima is useful because there lots of things in life that we want to maximize or minimize—profit, cost, risk, time, distance, and more.

Second-order differentiation—finding the derivative of the derivative—is important for finding maxima and minima. The fact that the derivative equals zero at some point really only guarantees that the curve is level at that point; it doesn't say whether it is the top of a peak, or the bottom of a valley, or a ledge halfway up a slope. The second derivative, however, helps distinguishes between these possibilities. Look at the point in Figure 11 where the lighter curve crosses the horizontal axis. The curve is decreasing there, so it has a negative slope—that is, its derivative, which is the second derivative of the darker curve, is negative. Now look up at the darker curve at that point: it is at a maximum. A curve's second derivative, then, tells us which way it is bending. If the second derivative is

negative, the curve is bending downward; if the second derivative is negative, the curve is bending upward. Curves bend downward at peaks (maxima) and upward at valleys (minima). This leads to the *second derivative test*: Where a function's first derivative is zero and its second derivative is negative, the function is at a maximum. Where its first derivative is zero and its second derivative is positive, the function is a maximum. If both the first and second derivatives are zero, the test fails and we have to investigate further, perhaps by investigating even higher-order derivatives.

Plotting the original curve and just looking at it is a perfectly good way to distinguish between maxima and minima when the function is simple, as in our examples, but in practice one may be looking at a function in two, three, or more dimensions, where visualization is difficult or impossible.

A Brief History of Discovery and Development

Calculus—or *the* calculus, as it is sometimes called—was invented in the 1600s independently and more or less simultaneously by the English physicist Newton and the German mathematician Leibniz. Actually, it is a slight exaggeration to say that either Leibniz or Newton invented calculus; both applied fresh insight to a mass of mathematical questions and tools that had been building up for centuries.

Mathematicians had been worrying about rates of change (what we now call derivatives) and the calculation of areas ever since the Greek mathematicians, such as Aristotle (384–322 B.C.).

But something new did come into being when the collection of techniques we now call "calculus" came together in the mid- to late 1600s. For the first time, mathematicians had a systematic way of finding derivatives and integrals, that is, rates of change on a curve and areas under a curve. Leibniz realized that to make calculus really useful, an easily understood system of notation would be needed—a way of writing down calculus that would do some of the work automatically. He achieved this by coming up with the integral sign, $\int$, and the notation for the derivative that we most commonly use today,

$$\frac{df}{dx}$$

In 1675, over 325 years ago, he was already writing calculus in his notebook using exactly the same notation we use today, such as

$$\int x^2\,dx = \frac{1}{3}x^3$$

This "Leibniz notation," as it is called, makes it possible for high-school and college students of normal mathematical ability to solve problems that baffled great mathematicians for centuries.

In the twentieth and twenty-first centuries, one of the most important practical developments in calculus—as in much of science and mathematics—has been the advent of the electronic digital computer, which was invented during World War II. Digital computers allow us to deal with equations that cannot be solved in "closed form," that is, reduced to a neat, algebraic expressions with the unknown variable on one side of the equals sign and all the known (or knowable) variables on the other. Equations that cannot be solved in closed form can arise even in simple problems. But using the computerized number-crunching techniques collectively referred to as "numerical methods," engineers, scientists, and others can today solve virtually any problem that can be stated using calculus, whether a closed-form solution can be found or not.

Interestingly, Leibniz and Newton still haunt the age of computerized calculus. One of the most commonly used numerical methods for solving equations was developed by Newton and is called "Newton's method," and Leibniz built one of the first mechanical calculating machines, a direct ancestor of the modern computer.

Real-life Applications

APPLICATIONS OF DERIVATIVES

Many direct applications of derivatives in real life are to situations of the sort called *min-max* problems or *extremum* problems, that is, situations where the goal is to find the minimum or maximum (the "extreme" values) of some physical, financial, or other quantity. Solving these problems requires derivatives. Derivatives are also used indirectly, as one of the many mathematical tools that engineers, scientists, and other math-using professionals need to solve their complex and many-layered problems. The following applications of calculus are simplified versions of real max-min problems, that is, they are direct applications of the concept of the derivative.

Maximizing Profits Making a profit is essential to the health and longevity of a business. And for any industrial

Credit for Calculus

Question: What do you get when you cross history with calculus? Answer: two famous mathematicians and nations arguing over who was first in its discovery.

Isaac Newton (1643–1727), staunch English Puritan and the England's champion of math and physics, developed the fundamental concepts of calculus in 1665 and 1666. He organized his ideas into a manuscript in late 1666 and showed it to a few other English mathematicians, but did not publish it. In 1672 to 1676, a German mathematician named Gottfried Leibniz (1646–1716), who started college at 15 and graduated at 17, worked privately on the same problems and came up with similar answers. Leibniz had not heard of Newton's work, and he developed notation and methods that were different from Newton's, but his ideas were essentially the same.

Leibniz first published his results in 1684 and 1686; Newton, in 1687. The math debate arose in the late 1690s, when followers of Newton began to accuse Leibniz of having stolen his calculus ideas from Newton. The fact that Leibniz had published first and Newton second might have made this impossible, but Newton and Leibniz had exchanged letters in 1676 and Leibniz had visited London in both 1673 and 1676, so it was not impossible that Leibniz had stolen Newton's ideas—merely untrue.

Newton and Leibniz actually invented calculus independently, not an uncommon event in science and mathematics. But sharing the accomplishment was not on anyone's agenda, especially in a question of national pride. Newton became so angry that he deleted all references to Leibniz's work from his scientific books (except insults). Newton and his followers publicly accused Leibniz of stealing. Leibniz asked the Royal Society of London, the major English scientific club or society of its day, to investigate this damning charge. Newton secretly stage-managed the society's investigation and Leibniz was found guilty.

Newton was buried in a cathedral with royal honors and thousands of mourners; Leibniz's funeral was attended only by his secretary. Leibniz's ultimate revenge, however, is that *his* calculus notation, not Newton's, is used today.

enterprise more complicated than a lemonade stand, it often calls for calculus.

Suppose that you are manufacturing x bottles of hair gel per day. It would be convenient if you could always make more profit just by cranking out more gel, but this does not work. You have to sell to make a profit, and you can only sell units as fast as your customers will buy them; making too many will result in unsold bottles and reduced profits. However, making too few units is no good either. Assuming that there are no factors affecting your profits other than how many bottles per day you manufacture—and that's a big assumption—how many bottles per day should you make?

Let's assume that you have discovered an equation that describes your daily profit, $p(x)$, as a function of how many bottles you make per day, x: $p(x) = -.03x^2 + 4x - 200$. Recall that you can find the minimum or maximum points of a function by finding where its derivative equals zero. The first step, then, is to take the derivative of the profit function. In this case, the derivative of $p(x)$ is $dp/dx = -.06x + 4$. To find where dp/dx equals zero, we solve the algebraic equation $-.06x + 4 = 0$ and find that the optimal number of bottles to manufacture per day is a whopping

$$\frac{4}{.06} = 66\,\frac{2}{3}$$

or, since it is presumably impossible to sell two thirds of a bottle of gel, 66 bottles.

No real-world business is this simple, but the principle is sound. Calculus, along with other branches of math (such as probability theory), is indeed used by financial analysts.

Storing Data on a Computer Disk Calculus has been used literally thousands of times in the design of every one of the electronic toys we take for granted including MP3 players, TV screens, computers, cell phones, etc. A simple example can be found inside the nearest computer.

For long-term information storage, every computer contains a "hard drive," the primary storage device that, if it crashes and you haven't made backups, will result in a loss of data. A hard drive contains a stack of thin discs

A compact disk is deteriorating along the edges and will no longer play properly. The larger the disk, the more susceptible it is to "disk rot." This and other factors limit disk size. Calculus is used to maximize the number of bits on the disk but the larger the disk the more susceptible it is to "disk rot." AP/WIDE WORLD PHOTOS. REPRODUCED BY PERMISSION.

coated with magnetic particles. A computer stores information in the form of binary digits ("bits" for short, 1s and 0s) on the surface of each disc by impressing or "writing" on it billions of tiny magnetic fields that point one way to signify "1", another way to signify "0." The bits are arranged in circular tracks, as shown in Figure 12.

To read information off the spinning disk, sensors glide back and forth between the edge of the disc and its center to place themselves over selected tracks. The track spins under the sensor, the bits are read off one by one at high speed, and within a few seconds, your favorite video game pops up. But that isn't the whole story. In designing a data-storage disc, if you should ever get the urge to do so, you will want to *optimize* the amount of data stored on the disk-that is, to store the most bits possible. How?

You might think that the way to do this would be to completely cover the disk's surface with tracks. Logical, but wrong, because there's one more real-life wrinkle, namely, that for the sake of keeping the read-write

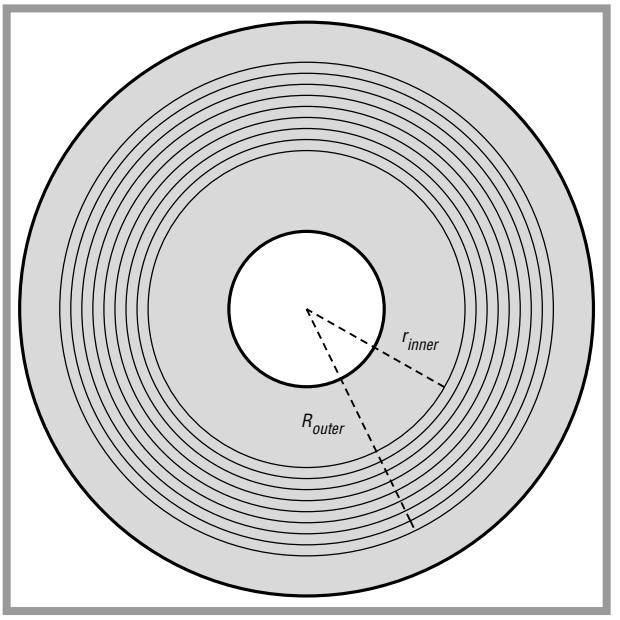


Figure 12: The shaded area is the readable-writeable part of the disc. Each circle represents a track. The radius of the innermost track is r_{inner}, that of outermost track is R_{outer}.

mechanism simpler (and therefore cheaper), every circular track has to have the same number of bits. So, by the formula for the circumference of a circle of radius r, $C = 2\pi r$, the bits are packed less densely on the outer tracks than on the inner tracks: same number of bits per track, more C to string them out on.

Say that the most bits you can pack onto each inch of track is b_i. Then, the most bits you can fit on a track of radius r is the length of the track times b_i, $2\pi r b_i$. So the smaller you make the radius of the innermost track, which we'll call r_{inner}, the fewer bits you can fit on it. But all the tracks on the disc must, as specified above, have the same number of bits, so if you make the innermost track too small, it will hold only a few bits, and so will all the other tracks, and you'll end up with an inefficient disc. On the other hand, if you make the innermost track too big, there won't be much room for additional tracks between the innermost track and the outer edge of the disc, and again your design will be inefficient. What to do?

Calculus to the rescue. Say that the most tracks that you can pack on the disk per radial inch (that is, going from the center toward the edge) is t_i. And let the radius of the outermost track on the disk be made as large as possible, the radius of the disk itself. Call this outer limit R. The number of tracks on the disk, then, is the radial distance between the innermost and outermost tracks times the number of tracks per inch that we can fit into that distance: total number of tracks equals $(R - r_{inner})t_i$. The total number of bits on the disk is the number of tracks times the number of bits per track, and the number of bits per track is limited by how many we can pack onto the smallest, innermost track, which from the previous paragraph we know to be $2\pi r_{inner} b_i$. So, writing the number of bits on the disc B as a function of r_{inner}, we have

$$B(r_{inner}) = \overbrace{2\pi r_{inner} b_i}^{\text{bits per track}} \overbrace{(R - r_{inner}) t_i}^{\text{tracks per disk}} = 2\pi b_i t_i (R r_{inner} - r_{inner}^2)$$

We want to maximize this function, $B(r_{inner})$, the number of bits on the disk. Taking the derivative using the rules found in standard calculus textbooks, we get

$$B'(r_{inner}) = 2\pi b_i t_i (R - 2r_{inner}).$$

To find where a function has maxima (or minima), we look for places where the derivative equals zero. So, setting B (r_{inner}) equal to zero and solving for r_{inner} using elementary algebra, we find that the value of r_{inner} that maximizes the number of bits on the disk is $r_{inner} = R/2$. Disk packed, case closed.

Lenses and Rainbows Light always takes the quickest possible path through whatever transparent materials it must travel through, a fact known as Fermat's principle. A physicist can write an equation that expresses the time taken by light to pass through an optical system (say, a series of lenses and reflectors). Taking the derivative of this function to find where its minimum point is shows what path that light will take through the system. Light-path minimization based on Fermat's principle is used in some computer programs for designing optical systems.

Designing for Strength In structures that must withstand strong forces, such as bridges and rockets, the force is never distributed evenly throughout the body of the object. Certain places, depending on the object's shape, will experience more than others. Engineers need to know where these points are and how big the maximum forces are that they experience; if the steel, stone, or other material from which the object is made isn't strong enough to withstand that maximum force, the bridge, rocket, or other object will fail. To find points of maximum force, designers describe force as a function and look for points where its first derivative equals zero. Some of these points may be minima, not maxima, but there are several ways—including the second derivative test—to tell which is which. Today, design of bridges, rockets, jet or car engines, and other complex structures is often done by making a mathematical image or model of the structure that can be stored in a computer. This model predicts force throughout the object as a function of space and time; wherever this function is at a maximum (first derivative zero, second derivative negative), if the predicted force is greater than the strength of the materials being used, the design must be changed.

Failure Prediction Derivatives can be used to guess when structures will break. Before breaking, many materials develop cracks. This causes them to emit brief noises that may not be audible to the human ear, but can be recorded by machines. One method of predicting structural breakdown is to use record the noises made by an object (e.g., a large, rotating metal shaft in a generator). The noises are counted by a computer, and their frequency (how fast they are happening) is recorded as a function of time. Software then calculates the first and second derivatives of this function and uses them to decide whether the noise frequency is increasing in a way that may indicate that a breakdown is going to happen soon. If the software detects a possible impending breakdown, it warns its human operators. Some scientists have proposed using the same method for predicting earthquakes, since an

A large section of the concrete roadway in the center span of the Tacoma Narrows bridge crashes into the Puget Sound in Tacoma, WA, on November 7, 1940. High winds caused the bridge to sway, undulate, and finally collapse under the strain. Engineers use calculus derivatives to estimate the forces bridges must withstand. AP/WIDE WORLD PHOTOS. REPRODUCED BY PERMISSION.

earthquake is essentially a sudden break in a large mass of material (the crust of the earth) that is preceded by a long period of strain and cracking.

Seeing Spectra Light waves vibrate at different rates or frequencies. One frequency or vibration-rate of light affects our eyes as the color blue; a slower vibration-rate (that is, a lower frequency) affects our eyes as the color red. There are also frequencies both too low and too high for our eyes to see. Most light is a mixture of many colors or frequencies, some of which are usually brighter than others. If we graph the brightness of the different colors in a beam of mixed light as a function of frequency, the resulting curve is called a spectrum. Often, a spectrum looks like a crowded row of tall, narrow peaks and valleys, all of different heights and depths. Spectra are used through science and engineering, for they contain information about the chemical composition of the objects giving off the light. Astronomers look at spectra to know what distant stars and planets are made of; chemists look at spectra in the laboratory to find out what chemicals their samples contain; and biologists and geologists look at spectra of the Earth's surface, as photographed by satellites, in order to map the composition and health of the earth's surface. Changes to the spectra of artificially pure beams of light (lasers) that have passed through the atmosphere are used to monitor pollution. First, second, and even higher-order derivatives of spectra are all used. In fact, in chemistry an entire sub-field, derivative spectroscopy, is devoted to the use of derivatives of spectra.

Derivatives make spectra more useful partly because they tend to *sharpen* the ups and downs in a function. If you compare a function to its first and second derivatives (as in Figures 2 and 11, for example), you'll see that where the original function is a gently curving slope, the first derivative has a peak and the second derivative has two sharper peaks. This sharpening effect can be used to clarify differences between the peaks and valleys in a spectrum, which makes it easier to tell what substances have reflected the light.

Sometimes derivatives of spectra are used even more directly. For example, the spectrum of light reflected from plants in tidal wetlands (as photographed by satellite) closely resembles, in part, the curve in Figure 2. The center of the rising part of the spectrum, where its derivative peaks (as shown in Figure 4), is called the "red edge" because here the spectrum drops to lower values for redder (lower-frequency) light. The exact location of the red edge indicates the amount of chlorophyll (a red-light-absorbing chemical) in the leaves reflecting the light, and so is an indicator of the health of the plants.

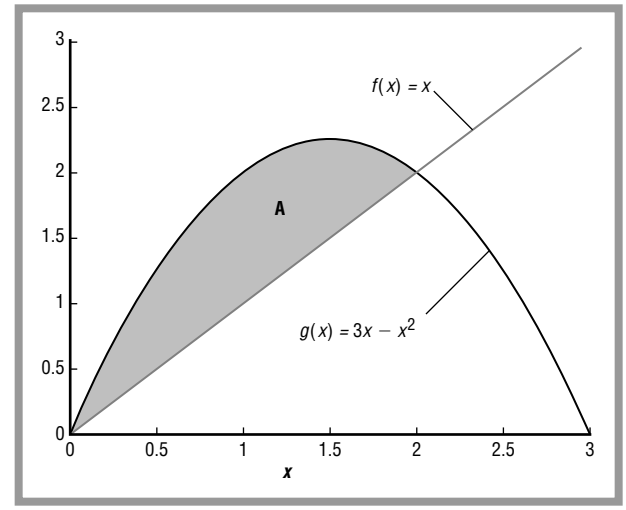


Figure 13: Difference of two functions. Integration can give an exact number value for the shaded area, A, between functions $g(x)$ and $f(x)$.

REAL-LIFE APPLICATIONS OF INTEGRALS

The Area Between Two Curves The integral of a function of one variable, as discussed earlier, is essentially the area under that function. What about the area between two functions, such as the shaded area in Figure 13?

With integration, determining the exact area of this oddly-shaped region is easy. The rule is this: To find the area between two functions $g(x)$ and $f(x)$, integrate the difference between them, $g(x) - f(x)$. In this case, $g(x) - f(x) = (3x - x^2) - x = 2x - x^2$. To find the area A between 0 and 2 in this example, we integrate this difference function between the two points, that is, we find the definite integral of $2x - x^2$ from 0 to 2. This can be written

$$A = \int_0^2 2x - x^2 dx$$

This expression can be evaluated using the rules of elementary integration given in all calculus textbooks. We find that A = 4/3.

Inertial Guidance Like all scientific knowledge, calculus can be applied not only to creation but to destruction. For example, the calculus-based concept of *inertial guidance* has been developed by missile-makers to a fine art.

The first ballistic missiles used in war, the V-2 rockets produced by Nazi Germany near the end of World War II (1939–1945), were fired at London from mainland Europe. They were intended as terror or "vengeance" weapons, and so only needed to explode somewhere over the city, not over particular military objectives; yet to hit a large city such as London at such a distance, a V-2 missile needed a guidance system, a way of knowing where it was at every moment so that it could steer toward its target. It was not practical to steer by the stars or Sun, because these are hard to observe from a missile in supersonic flight and would require complex calculations. Nor was it practical to steer by sending radio signals to the missiles, for without advanced radar (not yet available) controllers on the ground would be just as ignorant of the missile's location as the missile itself. Besides, the enemy might learn to fake or jam control signals, that is, drown them out with radio noise.

The solution was inertial guidance, which exploits the calculus fact that (a) the time derivative of position is velocity and (b) the time derivative of velocity is acceleration. By the fundamental theorem of calculus, which says that integration and derivative are opposites, we know that we can follow the trail backwards: the integral of acceleration is velocity, and the integral of velocity is position.

What designers need a missile to know is its position. But position is hard to measure directly. You have to look out the window, identify landmarks (if any happen to be visible), and do some fast geometry, likewise with velocity. But acceleration is easy to measure, because every part of an object accelerated by a force experiences that force. We've all felt the seat pushing against our backs in an

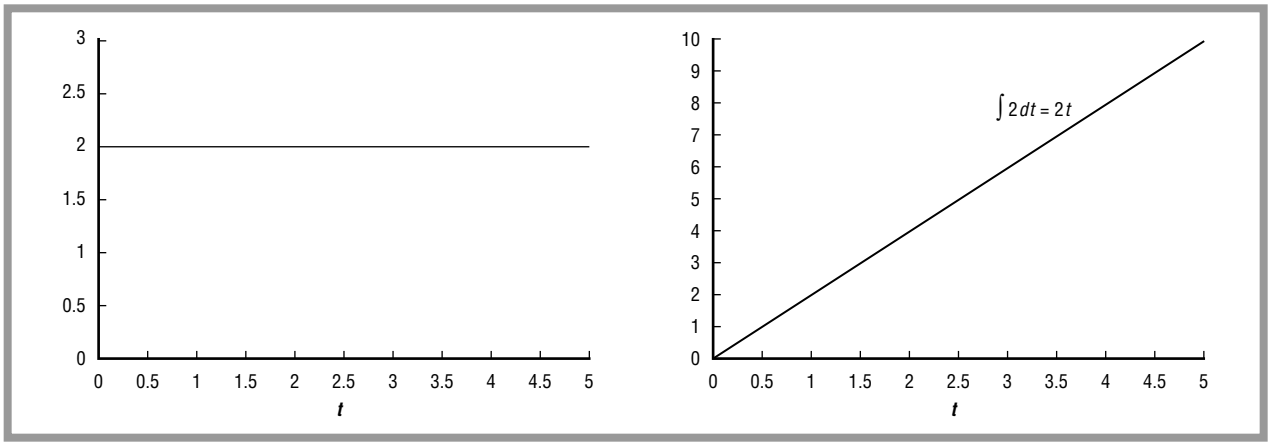


Figure 14: Left, a constant (the number 2) plotted as a function of *t*. Right, the definite integral of the number 2 from 0 to *t*. This shows that the integral of a constant is a linear function.

accelerating car or plane. In addition, unlike velocity or position, a force can be measured directly and locally, that is, without making observations of the outside world. Therefore, the V-2's engineers installed gyroscopes (spinning masses of metal) in their missile and used these to measure its accelerations. Lasers, semiconductors, and other gadgets have also been used since that time. Some are more expensive and accurate than others, but all do the same job: they measure accelerations. Any device that measures accelerations is called an accelerometer.

Thanks to its accelerometers, an inertial guidance system knows its own acceleration as a function of time. What does it do with this knowledge? Acceleration can be written as a function of time, $a(t)$. This function is known by direct measurement by accelerometers. The integral of $a(t)$ gives *velocity* as a function of time: $\int a(t)\, dt = v(t)$. And the integral of $v(t)$ gives *distance* as a function of time, which reveals one's position at any given moment: $\int v(t)\, dt = x(t)$. The real-world math of inertial guidance is of course more complex, but the principle is the same.

The bottom line for inertial guidance is that, given an accurate knowledge of its initial location and velocity, an inertial guidance system is completely independent of the outside world. It knows where it is, no matter where it goes, without ever having to make an observation.

The V-2 inertial guidance system was crude, but since World War II inertial guidance systems have become more accurate. In the early 1960s they were placed in the first intercontinental ballistic missiles (ICBMs), large missiles designed by the Soviet Union and the United States to fly to the far side of the planet in a few minutes and strike specific targets with nuclear warheads. They were also used in the Apollo moon rocket program and in nuclear submarines, which stay underwater for weeks or months without being able to make observations of the outside world. Inertial guidance systems are today not only in missiles but in tanks, some oceangoing ships, military helicopters, the Space Shuttle and other spacecraft, and commercial airliners making transoceanic journeys.

Calculus makes inertial guidance possible, but also, in a sense, limits its accuracy. The problem is called integration drift. Integration drift is a pesky result of the fact that small "biases" are, for various technical reasons, almost certain to creep into acceleration measurements. (A bias is a small, unknown number added to all your measurements, like .000000001 m/s^2.) Now, the integral of a constant (any ordinary number, like 2.0 or .000000001) is a linear function. That is, for any constant k, $\int k\, dt = kt$.

Figure 14 shows why this works. Plotted as a function of t, a constant k (e.g., the number 2) is a flat line: it never changes, it's always just k. But the area under that line grows steadily as one takes in more of the t axis, starting from any given point, such as 0.

What's more, the integral of a linear function is a quadratic function, that is, a function containing t^2 as its highest power of t:

$$\int kt\, dt = \frac{k}{2} t^2$$

This function is plotted in Figure 15, for $k = 2$.

Inertial guidance depends on measuring a physical variable (acceleration) and then, in essence, integrating

twice. Any bias in these acceleration measurements,any unwanted, constant number that adds itself to all the measurements, will result in a position error that increases in proportion to the square of time, just like the function in Figure 15. Notice how quickly the numbers can grow. After 1 second, a constant acceleration-measurement error of 2 m/s^2 produces a position error of 1 meter; after 3 seconds, an error of 9 meters; after 5 seconds, an error of 25 meters. These particular numbers are unrealistically large, but any degree of real-life quadratic error will eventually grow to unacceptable values. As a result, no inertial guidance system can go forever without taking an observation of the outside world to see where it really is. Increasingly, inertial guidance systems are designed to update themselves automatically by checking the global positioning system (GPS), a network of satellites that blanket the whole Earth with radio signals that can be used to determine a receiver's position accurately.

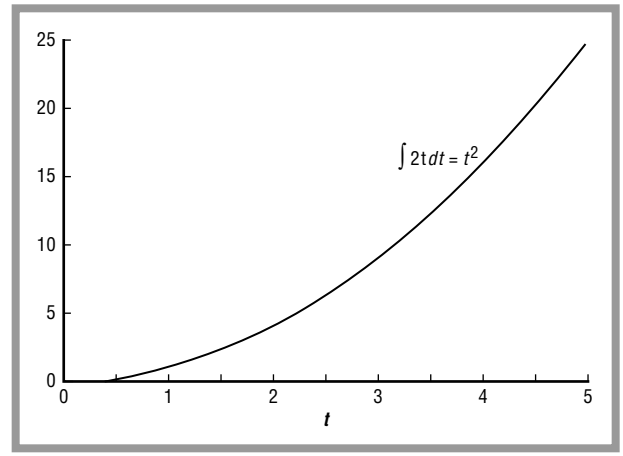


Figure 15: Showing that integrating a linear error (in this case, 2*t*) produces a quadratic error.

Today, inertial guidance systems have reached a fantastic degree of accuracy. A ballistic missile launched from a nuclear submarine, which may begin with a knowledge of its initial position and velocity that is several weeks old, can be launched from a still-submerged submarine by compressed air, burst through the surface of the water, ignite its rocket, fly blind to the far side of the planet, and explode its nuclear warhead within a few yards of its target.

Threading the Cosmic Needle On June 28, 2004, the Cassini space probe, a robot craft about the size of a small bus built jointly by the United States and Europe, arrived at Saturn after a seven-year journey. The plan was for it to be captured by Saturn's gravity and so become a permanent satellite, observing Saturn and its rings and moons for years to come. But to make the journey, Cassini had reached a speed of 53,000 miles per hour (85,295 km/h)—(many times faster than a rifle bullet), too fast to be captured by Saturn's gravity. At that speed it would swoop past Saturn and head out into deep space. Therefore, it was programmed to hit the brakes, to fire a rocket against its direction of travel as it approached its destination.

For objects moving in straight lines, changes in velocity can be calculated using basic algebra. Calculus is not needed. But Cassini was not moving in a straight line; it was falling through space on a curving path toward Saturn, being pulled more strongly by Saturn's gravity with every passing minute. To figure out when to start Cassini's rocket and how long to run it, the probe's human controllers on Earth had to use calculus. The effects of the important forces acting on Cassini—in particular, its own rocket motor and Saturn's gravity—had to be integrated over time. And the calculation—carried out using computers, not, for the most part, on paper—had to be extremely exact, or Cassini would be destroyed. For to get deep enough into Saturn's gravitational field, Cassini would have to steer right through a relatively narrow gap in Saturn's rings called the Cassini division (named after the same Italian astronomer as the probe itself), a navigational feat comparable to threading a cosmic needle. If it missed the gap, the Cassini craft would have been destroyed by collision with the rings.

Not all of NASA's navigational calculations have been correct: in 1998, a space probe crashed into Mars because of a math mistake. But in Cassini's case, the calculations were correct. Cassini passed the rings safely, was captured by Saturn's gravity, and began its orbits of Saturn.

Energy Payback Flip the switch and the light comes on. This simple act connects us directly to a vast system of electricity production worth many billions of dollars, with transmission wires marching across the countryside to immense yet delicately adjusted generating plants where the electricity is produced. Flipping a light-switch adds to the total demand for electricity that this system must meet.

The engineers who run our electricity production system are supposed to meet not only today's demand but the demand 10 and 20 years from now, so they try to forecast what that demand will look like. One mathematical prediction or "model" that has often been used to predict growth in demand for electricity is the exponential model. (See article on Exponents in this book.) This guesses that total demand for electricity will grow by a fixed percentage every year (e.g., 2%). That is, if electric demand is 100 units in the year 2025, it will be 102 units in 2026, and so forth, if the model is correct. Demand might grow because the population is growing and there

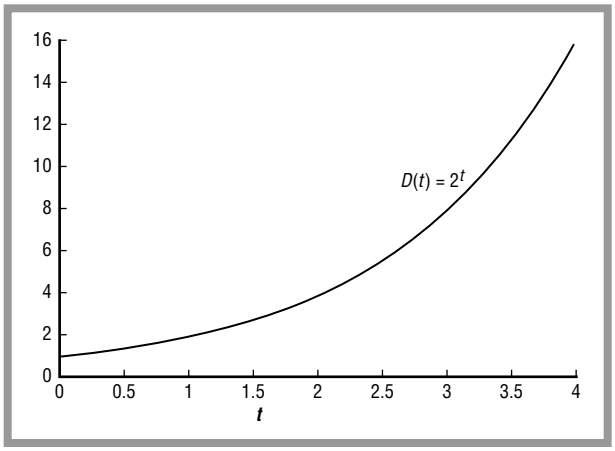


Figure 16: Exponentially growing demand for electricity modeled by $D(t) = 2^t$.

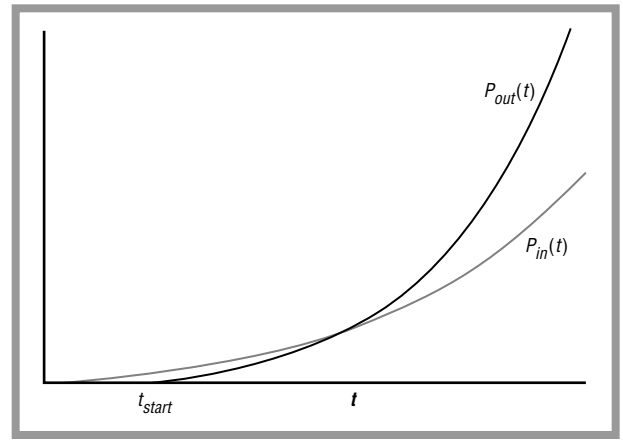


Figure 17: Total power output and power output of an exponential program for building power plants.

are more people using electricity, or because we are using more electricity to make things, or because we are wasting more electricity (by leaving our computers or lights on), or for some mixture of these reasons.

If we consider demand as a function of time, $D(t)$, then exponentially increasing demand can be written as $D(t) = D_0k^t$, where k is some constant (fixed number) and D_0, also a constant, is the value of $D(t)$ at $t = 0$. Any number to the 0th power is 1, so at time equals 0 we have $D(0) = D_0k^0 = D_0 \times 1 = D_0$. Exponential growth is depicted in Figure 16, with $D_0 = 1$ and $k = 2$.

Now, to meet new electricity demand, new electricity-production plants must be built. And to meet exponentially increasing demand, plants must be built at an exponentially increasing rate. So far, so clear—so why not build them? Because, even apart from the fact that on a finite planet no pattern of exponential growth can go on forever (it will eventually eat up the whole planet), we are confronted by a paradox: A program that builds power plants at an exponentially increasing rate will produce no energy for some time even after its first plants start generating electricity.

Why? Because it takes energy to *build* a power plant. That is, energy must be loaned from some other source. Before the plant can be a true energy producer, it has to pay back this energy debt. Furthermore, building power plants at an exponentially increasing rate requires an exponentially increasing amount of power. (Power is the rate of flow of energy, that is, its time derivative.)

Let's take a closer look at energy in and energy out. Call the power going into unfinished plants $P_{in}(t)$ and the power coming out of finished plants $P_{out}(t)$. We'll assume that the building program begins at time 0. At some later time, call it t_{start}, the first plant starts delivering electricity. For a large dam, coal-fired plant, or nuclear power plant, t_{start} might be 10 years. We'll also say that each finished plant produces more power than each unfinished plant uses.

Let the power output of each finished plant be p_{out} and the power input to each unfinished plant be p_{in}. Assume for that $p_{out} > p_{in}$. As plants are finished and start putting out power, the program as a whole will produce more and more power, but it will also consume more and more power, as construction of new plants grows exponentially too.

In real life, the total amount of power required by the building program increases by little jumps of size p_{in} (as construction of each new plant is begun), and the total amount of power produced increases by little jumps of size p_{out} (as each finished plant comes online). But here we'll pretend that both curves can be treated as smooth functions of time. We'll call the total power-investment curve $P_{in}(t)$ and the total power-output curve $P_{out}(t)$. Both curves climb exponentially, but the power-investment curve $P_{in}(t)$ starts to climb as soon as construction begins, namely at $t = 0$, and the power-output curve doesn't start to climb until t_{start}, when the first finished plant kicks in.

If $p_{out} > p_{in}$, then the power-output curve climbs more steeply than the power-investment curve. This situation is shown in Figure 17.

Notice in Figure 17 that although power output gets a late start, it soon catches up with power investment (where the curves cross) and surpasses it. But the program only becomes a net power producer when its summed power output exceeds its total power debt. When does this happen?

Before we look at the actual calculus, let's first emphasize that power is the time derivative of energy. That is, if a

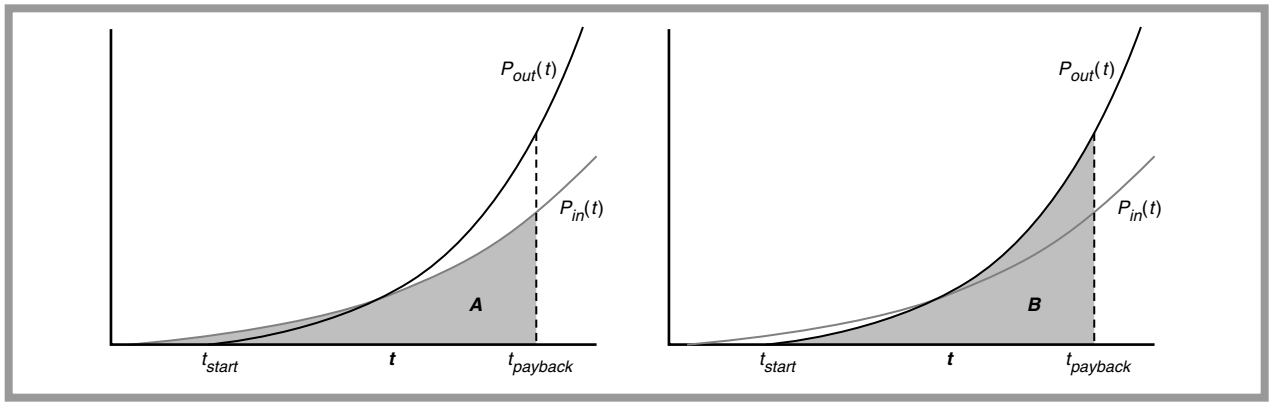


Figure 18: Area *A*, at left, and Area *B*, at right, correspond to total energy consumed and produced (respectively) in the time intervals shown. We want to find $t_{payback}$ such that $A = B$ to know when the building program pays off its energy debt.

curve records power (rate of energy transformation or "flow") as a function of time, then the area under that curve between any two times gives the energy supplied in that time. So to find out when our hypothetical program has first produced as much energy as it has consumed, we have to find that time (call it $t_{payback}$) when the area under the $P_{out}(t)$ curve equals the area under the $P_{in}(t)$ curve. That is, we have to find $t_{payback}$ such that area A equals area B in Figure 18.

Once again, calculus to the rescue. Area *A* (energy invested up to $t_{payback}$) can be written as the definite integral

$$A = \int_0^{t_{payback}} P_{in}(t)\,dt$$

Area *B* (energy produced up to $t_{payback}$) can be written

$$B = \int_{t_{start}}^{t_{payback}} P_{out}(t)\,dt$$

If A = B, then we just set these integrals equal to each other:

$$\int_0^{t_{payback}} P_{in}(t)dt = \int_{t_{start}}^{t_{payback}} P_{out}(t)dt$$

If we assign the simplest possible exponential forms to $P_{in}(t)$ and $P_{out}(t)$, this gives

$$\int_0^{t_{payback}} (e^{k_{in}t} - 1)dt = \int_{t_{start}}^{t_{payback}} (e^{k_{out}(t - t_{start})} - 1)dt$$

(Here the letter "e" stands for the constant 2.7182818 . . . , which is usually used in calculus for exponential functions.) Using the standard rules of integration found in calculus textbooks, this equation evaluates to

$$\frac{e^{k_{out}(t_{payback} - t_{start})}}{k_{out}} - \frac{1}{k_{out}} + t_{start} =$$

$$\frac{e^{k_{in}t_{payback}}}{k_{in}} - \frac{1}{k_{in}}$$

The only unknown in this equation is $t_{payback}$, so if we had particular numbers for the other variables, we could solve for $t_{payback}$ using a computer.

This general approach can be used to evaluate the realism of any proposed program to grow the electricity supply quickly. Simple-minded plans to rapidly build any kind of generating capacity—windmills, nuclear plants, coal plants, or other—in order to meet a projected energy shortage that is, say, 20 years away, may be worse than useless if $t_{payback}$ for the proposed program is 40 years!

Exponential functions have actually been used by government and industry analysts to predict growth in energy demand. A net-energy analysis such as that outlined above can reveal the long-term strengths or weaknesses of such predictions, and any energy solutions proposed to cope with such increases in demand. Exponential growth models can be relatively accurate over short periods of time and limited areas; fortunately, predictions of exponential growth in overall electricity demand have rarely turned out to be correct in the long term. This is mostly due to increases in user efficiency. For example, a typical refrigerator today uses about half as much electricity as a 10-year-old refrigerator but cools the same amount of food. Electricity costs money, so there is an ongoing economic pressure toward more efficient use.

Key Terms

Acceleration: A change of velocity (either in magnitude or direction).

Bit: The smallest unit of information storage in computers. A bit stores a 0 or a 1.

Derivative: The limiting value of the ratio expressing a change in a particular function that corresponds to a change in its independent variable. Also, the instantaneous rate of change or the slope of the line tangent to a graph of a function at a given point.

Differentiate: To determine the derivative or differential of a particular function.

Integral: The area under a curve.

Where to Learn More

Books

Edwards Jr., C.H. *The Historical Development of the Calculus.* New York: Springer-Verlag, 1979.

Price, John H. "Dynamic Energy Analysis and Nuclear Power," *Non-Nuclear Futures*, Lovins and Price, eds. Friends of the Earth International, 1975.

Sawyer, W.W. *What is Calculus About?* New York: Mathematical Association of America, 1961.

Straffin, Philip, ed. *Applications of Calculus, Vol. 3.* Washington, DC: Mathematical Association of America, 1997.

Web sites

"The Math Forum @ Drexel: Calculus" <http://mathforum.org/calculus/> (April 23, 2004).

"Visual Calculus: A collection of modules that can be used in the studying or teaching of calculus." <http://archives.math.utk.edu/visual.calculus/> (April 23, 2004).

Calendars

Overview

A feature in many kitchens is a calendar hanging on a wall. Typically, there will be all sorts of notations on the calendar, charting the various events and important points in the days and weeks. In electronic form, calendars have become a time and appointment tracking tool that serve to keep individuals organized both in their work and personal lives.

The prime function of a calendar is to organize time, over a short term and even extending far ahead into the future. This function hinges on math.

A Brief History of Discovery and Development

The need and desire to map the passage of time is something that has probably been with us ever since our prehistoric ancestors started to ponder the world around them. Gazing up at the sky would have made people aware of time. Days turn into nights and back into days. Then as now the moon waxed and waned in the night sky. With the advent of telescopes, the regular movement of some celestial bodies (like the planets in our solar system) was revealed. In more northern climates, seasonal variations in temperature and weather would have been apparent. All these things and more helped form the basis of the measurement of time.

Maintaining records of the passage of time has long been with us. Carvings and scratches in rocks, bones and sticks made by people some 20,000 years ago in present-day Europe are thought to be a form of calendar, to chart the appearance of the moon. Knowledge of when the nights would be bright with a full moon, or darker and better for sneaking up on game, would be beneficial for a hunter.

Five thousand years ago, the Sumerians who dwelled in what today is Iraq had a formal calendar. Their version divided the year into equal 30-day periods, each day into 12 equal periods (corresponding to 2 of our hours) and each of these daily periods into 30 parts (corresponding to about 4 of our minutes). Stonehenge, the jumble of stones assembled in southwest England over 4,000 years ago, was likely built at least in part to help chart universal events such as lunar eclipses and the passage of seasons.

Ancient Egyptians originally had a calendar based on the monthly cycle of the moon. However, they came to realize that a star (we know it as Sirius) appeared in the sky next to the sun every 365 days, at around the same time as the great river Nile flooded. This lead them,

Ancient Egyptians originally had a calendar based on the monthly cycle of the moon. This ancient Egyptian calendar is from Ptolemaic Alexandria, and show zodiac signs.
BETTMANN/CORBIS. REPRODUCED BY PERMISSION.

around 4236 B.C., to revamp their calendar, to one based on a 365-day cycle.

Elsewhere, the Mayan culture that flourished in Central America between 2600–1500 B.C. had 260 and 365-day calendars that were based on the Sun, Moon, and planet Venus. Portions of their calendars were sculpted into large calendar stones, which have survived to the present day.

The Julian calendar was introduced by Julius Caesar in 45 B.C. This version was a calendar based on the daily passage of the Sun across the sky. Each month was equal in length. Every fourth year a day was added to keep the calendar year in synch with the seasonal year. Today, principles of this calendar such as the fourth year added day and the January 1 beginning of a new year are still in use.

The year 45 B.C. is also known as the 'year of confusion' since Caesar inserted 90 days into the year to bring the calendar months back into synch with the seasons. It must have been a confusing year, indeed!

There are others examples of calendars. Indeed, even today there are approximately 40 different kinds of calendars in use.

Real-life Applications

As is evident that part of the math behind calendars is the segregation of time into units. For example, the Gregorian calendar that guides the days of many of us is another 365-day based design (except for every fourth year, the so-called leap year, when an extra day is added on to the month of February). Each 365-day period is divided into collections of days, usually 30 and 31, except for the 28 or 29 days of February, that are called months. In turn, the days in each month are organized into groups of seven, each of which represents a week. Fine-tuning things further, each day is divided into the 24 equal splits of time called hours and each hour into 60 minutes. Further divisions are possible.

The math at the heart of calendars organizes and at least gives the sense of controlling time. So, the math connects people to the world and even to the universe. It is not surprising that calendars assume such central and even sacred importance to societies throughout recorded history and even back into the mists of time.

The daily division of time in many calendars is based on astrological events. One is the daily cycle that results from the rotation of the Earth on its axis. Because the Earth is moving, relative to the sun, any particular portion of the globe will light and dark periods. The Gregorian and Julian calendars are solar calendars.

The monthly calendar cycle is based on the revolution of the Moon around the Earth. This is a lunar calendar. An example of a lunar-based calendar is the Islamic calendar. Because the phases of the moon do not match up with the months of the year, the Gregorian and Islamic calendars do not 'match up.'

Calendars also track the length of the four seasons. The basis of seasons is also astrological. The Earth's north-south axis is not oriented at 90 degrees to the Sun. Rather, the axis is tilted, with the North Pole being further away from the Sun than the South Pole. The result is that, as the Earth revolves around the Sun, the sunlight is more intense over certain regions of the planet at different times of year.

This tropical year is incorporated into a third type of calendar that is a blend of the solar and lunar calendars. The Hebrew and Chinese calendars are examples of this blend, which is called a lunisolar calendar. A lunisolar calendar has a sequence of months that are based on the cycle of the moon. But, every few years a month is added in, to bring the calendar back in synch with the tropical year.

LEAP YEAR

As mentioned earlier, the Julian and Gregorian calendars have some years that are one day longer. In the Gregorian calendar these leap years are 366-days long, rather than the usual 365-day year. The determination of when a leap year occurs is a straightforward mathematical process. Every year that can be divided evenly by 4 is a leap year, except for years that can also be divided evenly by 100. The latter, those years that mark the end of a century, can be leap years, but only when they can be divided evenly by 400.

Using these rules, the year 2000 is a leap year, since it can be divided exactly by 4 (to yield 500), by 100 (to yield 20) and by 400 (to yield 5). But the year 1900 is not a leap year. This is because it can be divided evenly by 4 (to yield 475) and by 100 (to yield 19), but cannot be divided exactly by 400 (1900/400 = 4.75). The year 2100 is also not a leap year.

The 400-year cycle of the Gregorian calendar comprises 146,097 days. Dividing the number of days by the number of years results in 365.24 (just about the number of days in each year). Multiplying this number by 2 or 3 does not produce a whole number. But, when 365.24 is multiplied by 4, the result is very close to a whole number. This is the basis of the 4-year cycle of leap years.

MATHEMATICAL ORIGIN OF THE GREGORIAN CALENDAR

The Gregorian calendar was devised to recalculate the dates of Easter. Centuries ago, the March 21 date of Easter coincided with the spring equinox; one of two days each year when the length of daytime and nighttime are the same at 12 hours. But, by the thirteenth century, people became aware that Easter was falling earlier in the month than the equinox. Popes Pius V and Gregory XIII worked on readjusting things. The solution implemented by Gregory was to delete October 5 through October 14, 1582, from the calendar. In that year, October 4 was followed by October 15, which put occurrence of the spring equinox back around March 21.

MATH AND THE ISLAMIC AND CHINESE CALENDARS

In the Islamic calendar, the months correspond to the lunar cycle. Twelve lunar cycles comprises a period in the Gregorian calendar equivalent to about 33 years.

The Chinese calendar is a lunisolar calendar whose months depend on the positions of the Sun and Moon. The pattern of 29- or 30-day months forms the basis for a 60-year cycle of names. A year name consists of a name from a group of celestial names and terrestrial names. The latter are names of animals and is the basis of Chinese years such as 'Year of the Rat' and 'Year of the Pig'.

For all the different calendar systems that have and still exist, several fundamentals are common. One is the intent of a calendar to organize time. The other is the vital relationship of math to the structure of the calendar. As in many other aspects of life, calendars are all about real-life math.

Where to Learn More

Books

Bourgoing, J.D. *Discoveries: The Calendar History, Lore, and Legend*. New York: Harry N. Abrams, 2001.

Judge, M. *The Dance of Time: The Origins of the Calendar – A Miscellany of History and Myth*. New York: Arcade Publishing, 2004.

Richards, E.G. *Mapping Time: The Calendar and Its History*. New York: Oxford University Press, 2000.

Web sites

Doggett, L.E. "Calendars." <http://astro.nmsu.edu/~lhuber/leaplist.html> (November 3, 2004).

Cartography

Overview

Cartography—the making of maps and charts—relies on mathematical principles to produce accurate representations of features or attributes distributed in space and time. The primary uses of mathematics in cartography are to accurately transform the spatial relationships among features on a curved surface onto a plane such as a piece of paper or a computer monitor and to determine the precise locations of features. Maps can depict physical features such as cities and roads, the type of bedrock exposed at the surface, or the elevation of the ground surface. They can also depict non-physical attributes such as the likelihood of damage during a strong earthquake or the average income of residents. The word chart is usually restricted to highly specialized maps showing coastlines, water depths, navigational aids such as buoys, and navigational hazards such as reefs in great detail.

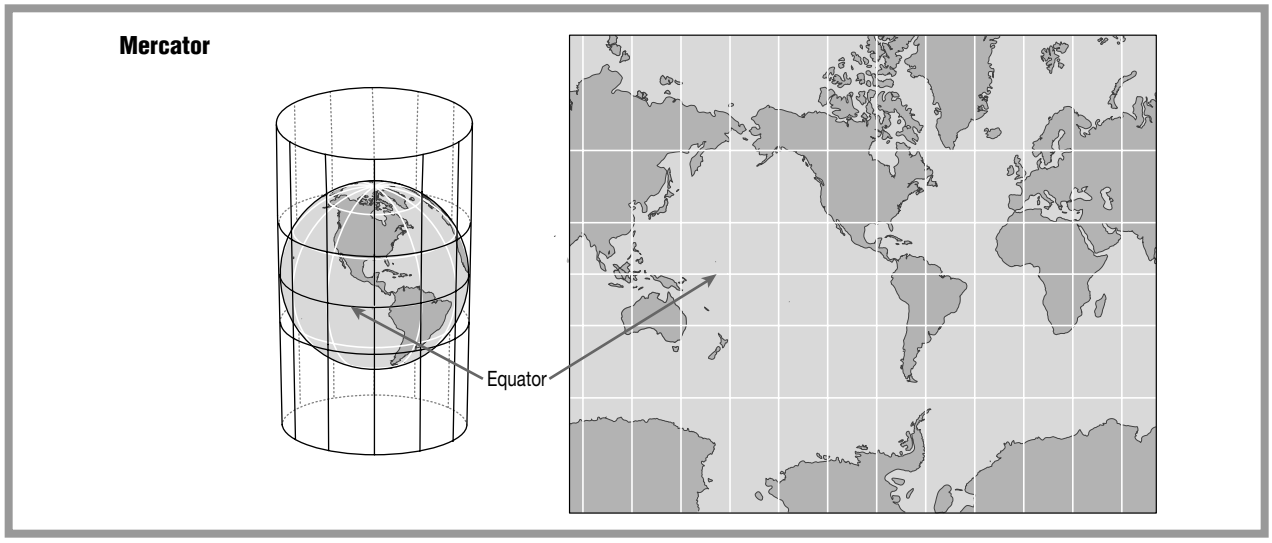
Fundamental Mathematical Concepts and Terms

SCALE

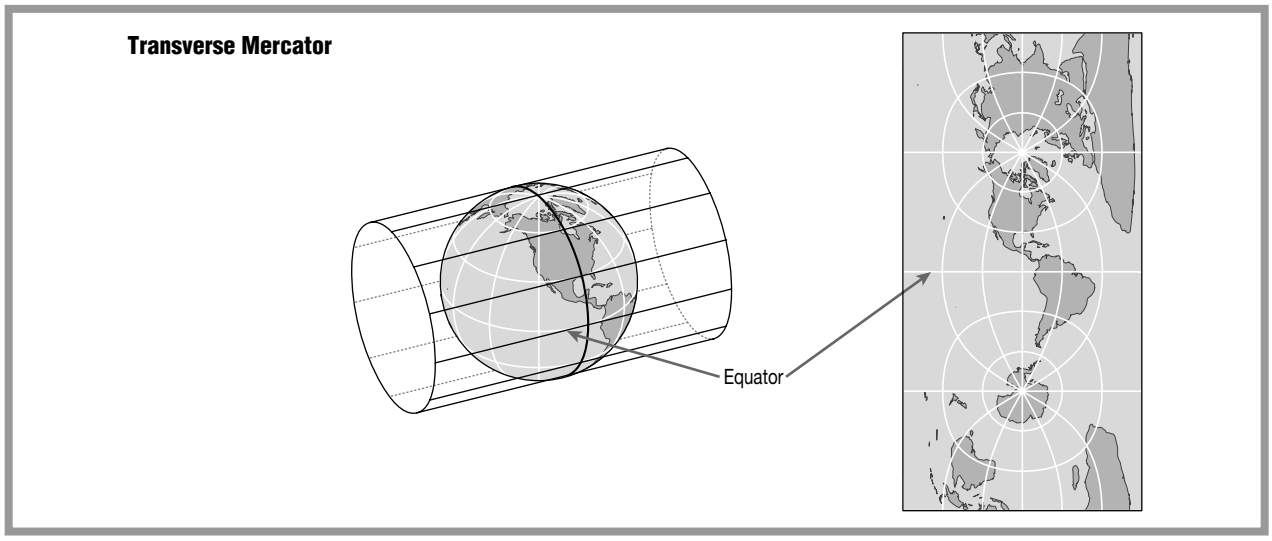
The scale of a map is the mathematical relationship between real distances and those shown on the map. If two buildings located 1 mi (1.6 km) apart are shown 1 in (2.5 cm) apart on a map, the map scale is 1 in : 1 mi or, using consistent units, 1:63,360. If the same two buildings are located 0.1 in (0.25 cm) apart on the map, then the scale becomes 1:633,600. Scale can also be written in a fractional form, for example 1/100,000, or shown graphically by a scale bar printed on the map. A scale of 1:100,000 is said to be larger than a scale of 1:1,000,000 because the fraction 1/100,000 is larger than 1/1,000,000. The concept of large and small scale can sometimes be confusing because a small scale map will cover a larger area than a large scale map of the same size. Large scale maps, however, show more detail than small scale maps covering the same area. Although there is no universally accepted definition of the difference between large, intermediate, and small scale maps, those with scales larger than 1:25,000 are usually considered large scale and those with scales less than 1:250,000 are usually considered small scale. Maps with scales between 1:25,000 and 1:250,000 are generally considered to be intermediate scale. To eliminate the possibility of confusion, it is always best to specify the scale numerically rather than just qualitatively using words like large or small.

MAP PROJECTION

One of the most difficult problems facing cartographers is the development of methods to transfer Earth's



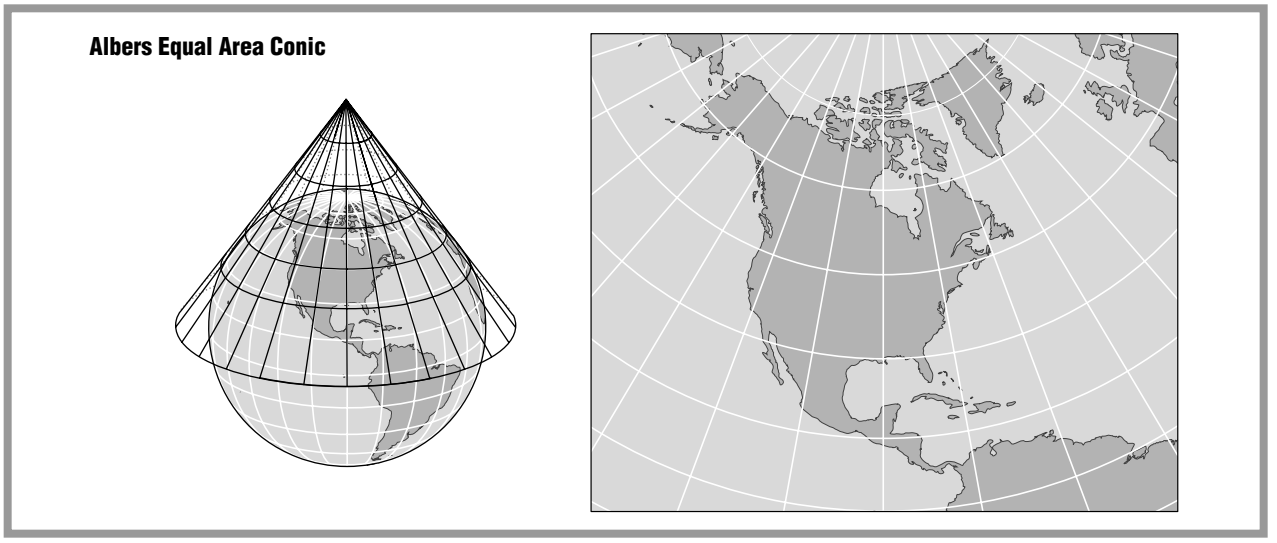
A Mercator map. MAP BY XNR PRODUCTIONS, INC. THE GALE GROUP.



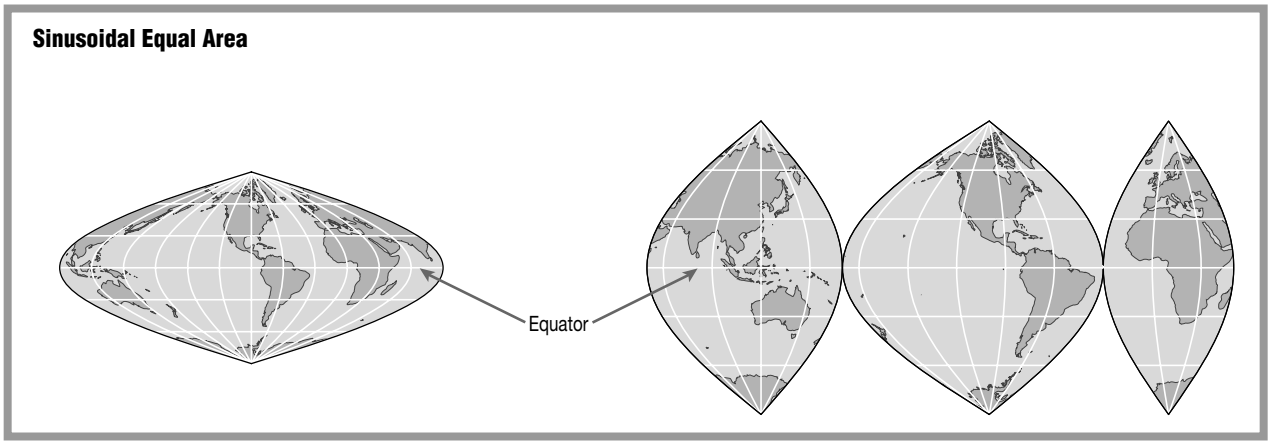
A transverse mercator map. MAP BY XNR PRODUCTIONS, INC. THE GALE GROUP.

spherical shape onto flat pieces of paper. This is done using an application of geometry known as map projection. Cartographers have developed dozens of different projections over the years, each of which has its advantages and disadvantages. Regardless of the projection used, however, all maps produced on flat surfaces will include some distortion.

The Mercator projection, invented by the cartographer Gerardus Mercator in 1569, is useful for navigation because a straight line drawn on a Mercator map represents a straight line on Earth's surface. Because Earth is spherical, however, the straight line does not necessarily represent the shortest distance between two points. Mercator projections are projections of the spherical Earth onto a vertical cylinder tangent to the Equator. Therefore, a Mercator projected map will occupy an unbroken rectangle representing an unrolled cylinder. Its primary disadvantage is that the Mercator projection distorts areas and shapes to a degree that increases as one moves away from the Equator. Therefore, land masses at mid- to high latitudes appear disproportionately large. A variation on the Mercator projection, the transverse Mercator projection, was created by Johann Heinrich Lambert in 1772. It is a projection of the spherical Earth onto a horizontal cylinder. The Mercator projection was brought into the space age during the 1970s, when cartographers



An Albers equal area conic map. MAP BY XNR PRODUCTIONS, INC. THE GALE GROUP.



A sinusoidal equal area map. MAP BY XNR PRODUCTIONS, INC. THE GALE GROUP.

A.P. Colvocoresses, J.P. Snyder, and J.L. Junkins invented the space oblique Mercator projection in order to project images obtained from Landsat satellites orbiting Earth.

As the name implies, conic projections use a cone that is tangent to Earth's surface rather than a cylinder. One kind of conic projection, the Albers equal area conic projection, is used by the U.S. Geological Survey for maps of the conterminous United States because it is well-suited for areas that have a large east-west extent. Another conic projection, the Lambert conformal conic projection, is also well-suited to areas with large east-west extents and is used for many maps of the United States.

Sinusoidal equal area projections, which have been in use since 1570, avoid the distortion of Mercator projections and are often used for maps in which it is important to compare the sizes or shapes of features in different parts of the world. An example of this would be a map showing the distribution of oil fields around the globe. A Mercator projected map would exaggerate the sizes of oilfields far from the Equator, but a sinusoidal equal area projection does not. A disadvantage of the sinusoidal equal area projection is that its projected shape is a series of lozenges or pods rather than a simple rectangle.

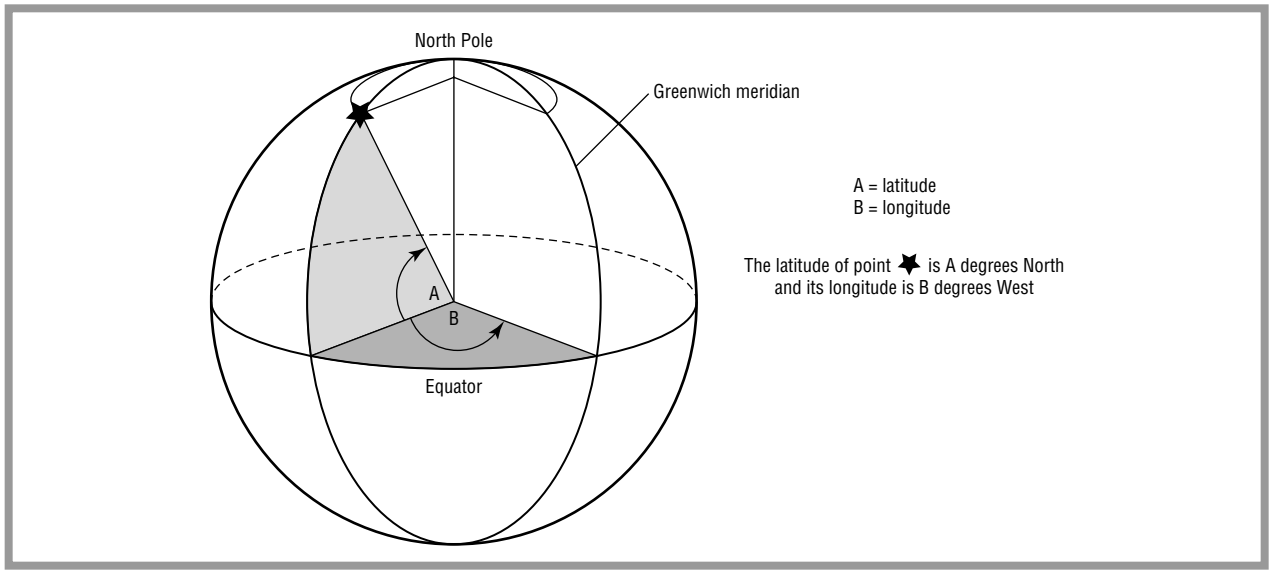


Figure 1.

COORDINATE SYSTEMS

Coordinate systems provide a way to record the position of a feature on the ground or on a map. The most widely known geographic coordinate system consists of lines of latitude and longitude on a sphere, which are measured as angles north or south of the Equator for latitude and east or west of the Greenwich (or Prime) Meridian for longitude (see Figure 1).

Angles of latitude range from 0 at the Equator to 90 degrees at the North and South Poles, and angles of longitude can range from 0 to 360 degrees. It is important to specify whether latitude is north or south of the Equator and whether longitude is east or west of the Greenwich Meridian. Lines of latitude are parallel to each other and each degree of latitude is equivalent to about 70 mi (112 km). Lines of longitude, which are known as meridians, converge at the North and South Poles and the distance between them decreases away from the Equator. The distance between meridians is about 70 mi (112 km) at the equator and decreases to zero at the two poles. Because there is such a large distance between lines of latitude and longitude, it is convenient to subdivide each degree into smaller parts. Degrees (°) have traditionally been divided into 60 minutes (′) of latitude or longitude, and minutes into 60 seconds (″) of latitude or longitude. For example, the location of the Seattle-Tacoma International Airport is 47° 26′ 56″ North latitude and 122° 18′ 34″ using degrees, minutes, and seconds. Because it can be difficult to perform arithmetic with latitude and longitude values given in degrees, minutes, and seconds, latitude and longitude can also be specified in decimal degrees. The location of the Seattle-Tacoma airport in decimal degrees is 47.45° North latitude and 122.31 ° West longitude.

Latitude and longitude are well suited for locating points on spheres and global navigation, but can be inconvenient to use for small areas. Therefore, other coordinate systems have been developed over the years. One of these is the Universal Transverse Mercator (UTM) grid system, which divides the globe into 60 zones each 6° of longitude wide and divided into North and South halves. Each zone has its own Transverse Mercator projection, which allows the curved surface of Earth to be accurately projected onto a flat map with grid lines that are parallel and perpendicular to each other. Because Earth is covered by 60 projections, distortion within any one of the zones is minimal. The UTM coordinates of a point are given by specifying the zone number, an east-west distance known as the easting, and the distance north or south of the Equator, which is known as the northing. UTM positions are always given in meters and never in feet or miles. Using UTM coordinates, the location of Seattle-Tacoma International Airport is Zone 10 N, 552,058 E, 5,255,280 N. One disadvantage of UTM coordinates is that is that they cannot be extended beyond zone boundaries. Therefore, latitude and longitude remain the standard for global mapping and navigation.

Each state within the United States also has its own coordinate system, known as a State Plane Coordinate System, that is used by surveyors and government agencies. State plane coordinate systems give distances east and north of a specified point in each state, can can be divided into parts for large states. Unlike UTM coordinates, state plane

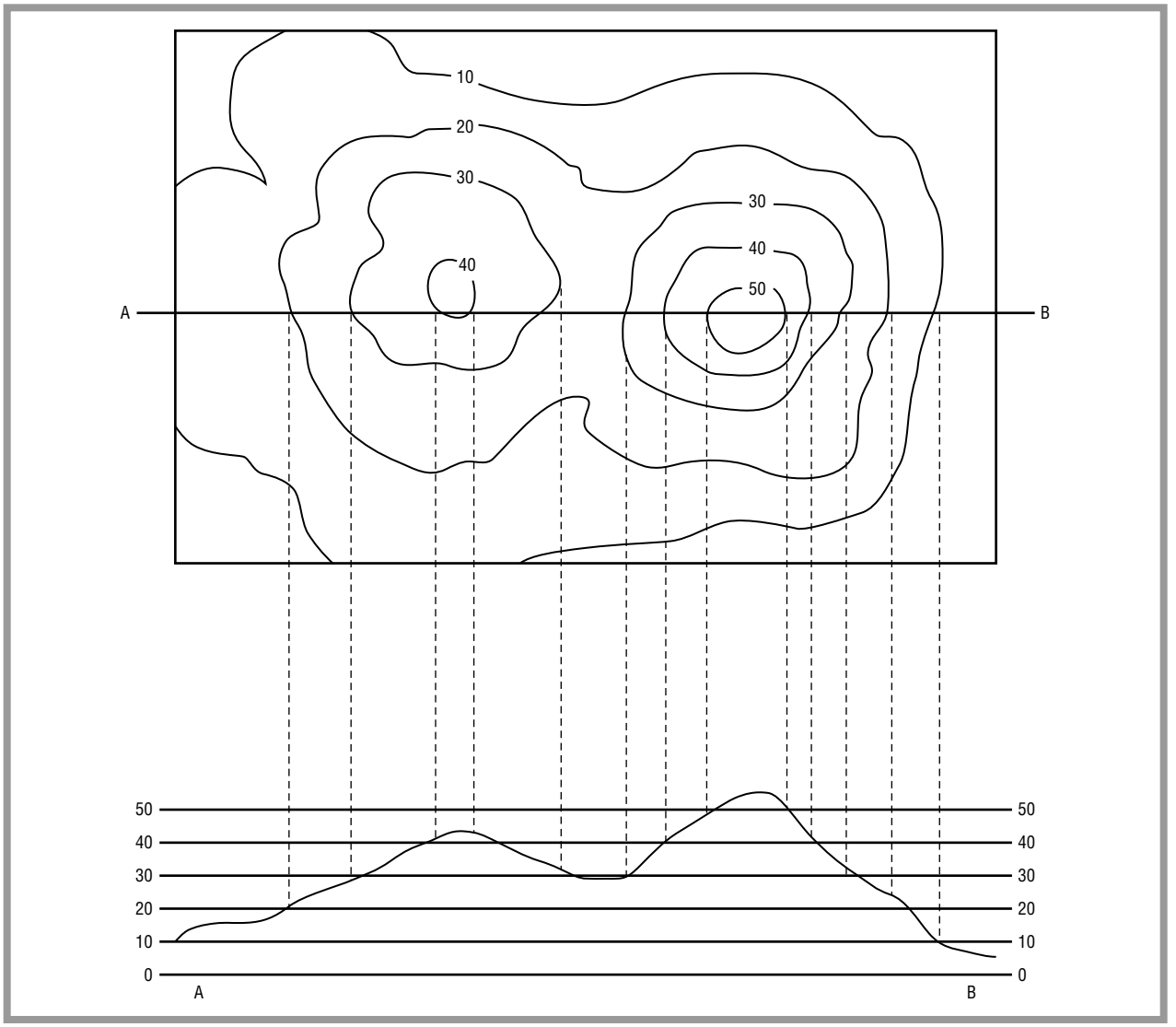


Figure 2.

coordinates can be given in either meters or feet (although it is important to specify which units are being used).

TOPOGRAPHIC MAPS

Topographic maps (an example is shown in Figure 2) are specialized maps that show the elevation of Earth's surface using contour lines, which connect points of equal elevation, or shading. They are especially useful for route finding, search and rescue operations, construction site selection, and scientific studies because of their accurate representation of landforms such as hills and valleys. Skilled map readers can easily interpret patterns of contour lines and visualize the landscape. Topography can also be represented using digital elevation models, which are computer files containing the elevation of Earth's surface at thousands or even millions of known points. Digital elevation models can be used to create contour maps, shaded relief maps, or three-dimensional surfaces that are useful for many applications.

A Brief History of Discovery and Development

Maps have been important in warfare, agriculture, trade, and the growth of civilizations for thousands of years. The oldest map discovered to date is a small clay tablet unearthed in 1930 at an archeological excavation in

Iraq, at the ancient city of Ga-Sur about 200 mi (320 km) north of present-day Baghdad. The tablet is thought to date from the time of Sargon of Akkad (2334–2279 B.C.) and shows two hilly areas on either side of a stream. More than two millenia passed before the Greek mathematician Eratosthenes (276–194 B.C.) estimated the circumference of Earth by measuring the lengths of noon shadows cast at two distant points at midsummer. The Greek astronomer Ptolemy (ca. 100–170, exact dates unknown) produced the first maps showing Earth as a sphere and included lines of latitude and longitude, but his maps were not widely used until the fifteenth century. Scientific cartography languished through the Middle Ages but blossomed again in the Renaissance and the Age of Exploration, when accurate maps became so important that they were considered to be military, diplomatic, and commercial secrets. It was at this time that Mercator developed his projection and produced a widely known map of the world. Map production technology continued to advance through the centuries, spurred on by events such as the introduction of aerial photography in the early twentieth century, satellite-based remote sensing in the 1960s and 1970s, and the widespread use of geographic information systems (GIS) software in the 1980s and 1990s.

Real-life Applications

GPS NAVIGATION

Maps are an important part of global positioning system (GPS) navigation systems installed in personal automobiles, commercial vehicles, aircraft, and ships. GPS receivers obtain signals from a system of satellites and, in the best of circumstances, can calculate locations with an accuracy of a few feet. In order to be useful for aviation or marine navigation, the location provided by the GPS receiver must be combined with an accurate map showing navigational hazards such as mountain ranges or submerged reefs.

GIS-BASED SITE SELECTION

GIS software allows map users to combine different kinds of maps in order to select the best locations for everything from new stores to hazardous waste storage facilities. Maps showing topography, transportation routes, population, natural hazards such as flood plains, and many other factors can be combined and sites selected using sophisticated mathematical algorithms that weigh the importance of information contain on each map or determine the most economical route between two locations. Maps and GIS systems are also used to understand the spread of diseases, evaluate patterns of criminal activity, and distribute aid in the wake of natural disasters such as floods, hurricanes, and earthquakes.

Key Terms

Cartographic projection: A geometric transfer of patterns, shapes, and locations from a spherical globe to a flat surface.

Scale: The ratio of the size of an object to the size of its representation.

NATURAL RESOURCES EVALUATION AND PROTECTION

Scientists and engineers use maps on a daily basis to record the distribution of natural resources ranging from ore deposits to endangered species habitat, to depict the locations of hazards such as landslides or earthquake faults, and to regulate activities such as commercial logging and urban growth. Some of their cartographic products include geologic maps showing the distribution of rock types, soil survey maps useful for agriculture and land use planning, and land cover maps that can be used to help assess the likelihood for erosion during heavy rainstorms.

Potential Applications

Maps will remain vital tools as long as humans continue to travel and explore. In the future, maps may help to guide the exploration of nearby planets and moons. Maps and GIS software will continue to integrated into cellular phones and portable computers used in day-to-day activities such as shopping and vacation travel. Computer mapping software and databases will allow users to create maps custom tailored to their unique needs.

Where to Learn More

Books

Crane, Nicholas. *Mercator: The Man Who Mapped the Planet.* New York: Henry Hold and Company, 2003.

Thrower, N.J.W. *Maps and Civilization: Cartography in Culture and Society*, 2nd ed. Chicago: University of Chicago Press, 1999.

Wilfor, J.N. *The Mapmakers (Revised)*. New York: Knopf, 2000.

Web sites

Aber, J.S. "Brief History of Maps and Cartography". 2004. <http://academic.emporia.edu/aberjame/map/h_map/h_map.htm> (February 12, 2005).

Cartography Associates. "David Rumsey Map Collection". 2003. <http://www.davidrumsey.com/> (February 12, 2005).

U.S. Department of the Interior. "National Atlas Home Page". January 26, 2005. <http://nationalatlas.gov/> (February 12, 2005).

U.S. Geological Survey. "Map Projection Publications", Fact Sheet 087-99, May 1999. February 19, 2004. <http://erg.usgs.gov/isb/pubs/factsheets/fs08799.html> (February 12, 2005).

U.S. Geological Survey. "A Tapestry of Time and Terrain: The Union of Two Maps—Geology and Topography." March 29, 2002. <http://tapestry.usgs.gov/two/two.html> (February 12, 2005).

Woodward, David. "The History of Cartography." January 11, 2005. February 19, 2004. <http://feature.geography.wisc.edu/histcart/> (February 12, 2005).

Charts

Overview

Charts are a graphical representation of quantitative data, usually laid out in two-dimensional form. A chart is a picture of related data, or, with line charts, a picture of an equation.

Charts and graphs are used extensively to help people both communicate numerical information and understand the relationship between different data sets. Charts are a way to view numbers and/or math, in picture form.

Charting grids are called a planes because of their two-dimensional qualities, usually represented in x-axis and y-axis forms, where the x axis is the horizontal plane and the y axis is the vertical plane. Each point, or coordinate, represents the x-value location relative to the y-value location. Charts are also commonly called graphs because the exercise of plotting these x and y coordinates is known as graphing. This mapping process follows a diagram where the variation of a dependent variable is in comparison with another, independent variable.

Fundamental Mathematical Concepts and Terms

There are three basic chart formations: line charts, column charts, and pie charts.

Line charts are pictures plotted, connected dots, often representing equations. Column charts are pictures of data clusters, while pie charts are pictures of percentages.

BASIC CHARTS

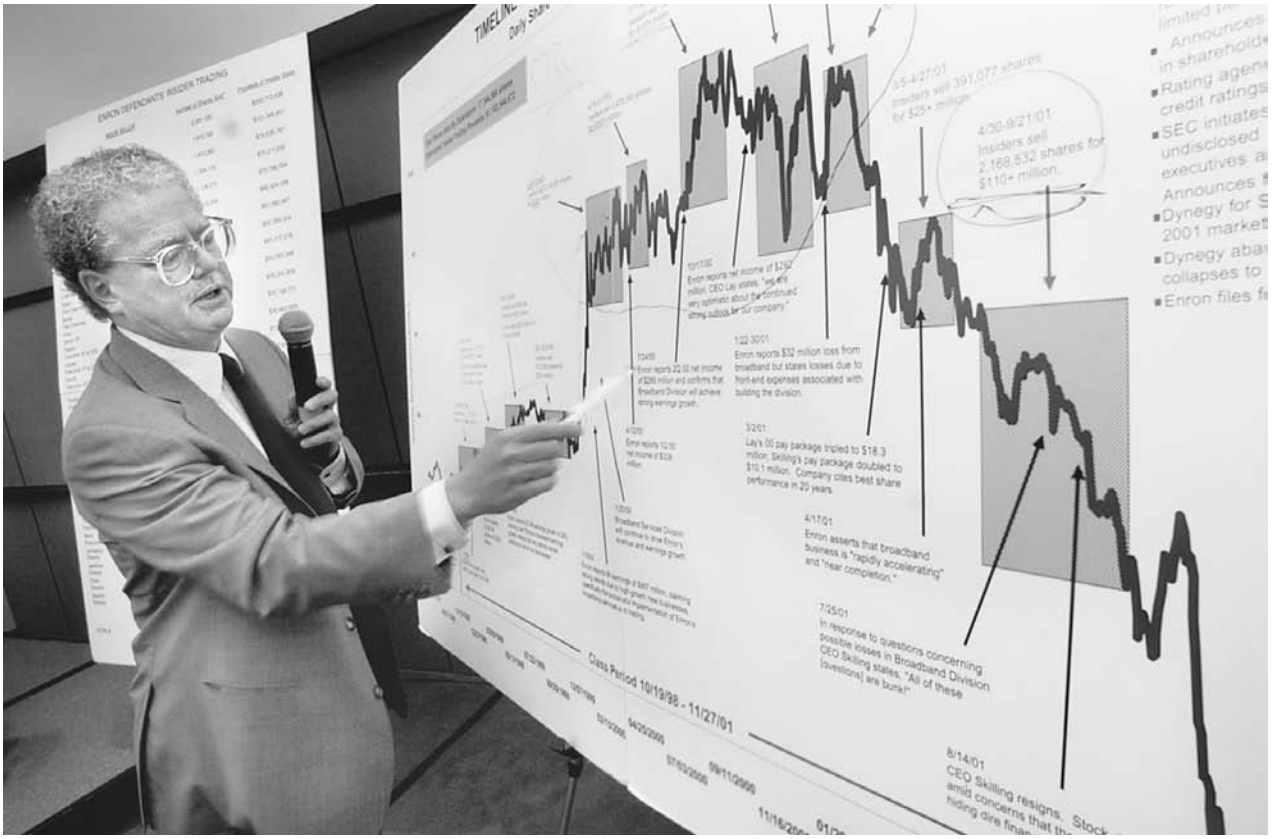
Although the line chart and the column chart are similar, each has unique characteristics that help the user better understand the data. The pie chart represents parts of a whole (percentages) and shows how parts of a data set are combined to create a complete picture.

There are many variations of these three basic charts, but the motivation between choosing which chart to use depends on how to communicate the information.

LINE CHARTS

Line charts are commonly used to graph time horizon data such as stock prices, economic data, or sales of specific companies. Many different types of data can be plotted in a line chart but the main benefit is to show how a data set trends, usually over a time horizon. Figure 1 shows a line chart in which the line goes up or down.

A line chart is a series of connected dots. Each dot represents a coordinate value of independent and dependent



Attorney William Lerach, representing Amalgamated Bank, uses a timeline chart of the Enron Corp. daily share price at a New York news conference in 2001. AP/WIDE WORLD PHOTOS. REPRODUCED BY PERMISSION.

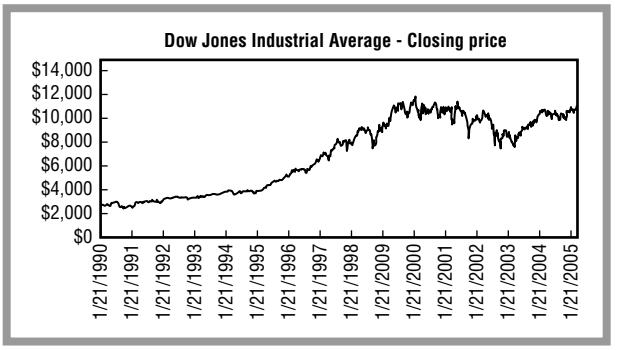


Figure 1.

variables. Usually the independent variable is plotted on the x axis (horizontal), and the dependent variable is plotted on the y axis. Each coordinate of the x value is plotted in relation to its y value. The dots are then connected to create a line.

When plotting changes over time, the x axis usually represents the time, the independent variable, and the y axis represents the dependent variable, the data values. With all line charts, the order of the values on both the x axis and the y axis is relational.

In line charts, the slope of the line between two points is easily recognized simply by observation. No calculations are necessary to get a good idea of the relative change from one segment to the next. Another quality about line charts is that it is easy to see if the values are increasing or decreasing along different segments. This change is often referred to as volatility.

Multiple-line charts can be plotted on one graph to show the relationship of different data sets with similar parameters relative to each other. A good example of multiple-line charts is seen daily when looking at the price changes of two similar stocks trading on the New York Stock Exchange. Figure 2 shows how to plot both lines on the same graph and quickly see which stock yields the better return for its investors.

In a double-axis line chart, one or more lines is plotted on the left axis, and one or more lines is plotted on the right y axis. This is useful when looking at the relative changes of different data that have similar independent

variables (e.g., time horizon), but the data does not have similar dependent variables. Using the stock price example, Figure 3 shows how to compare the price change of two different companies over the same time horizon, but they have very different stock prices. One set of data is plotted on the left y axis, and the other on the right y axis.

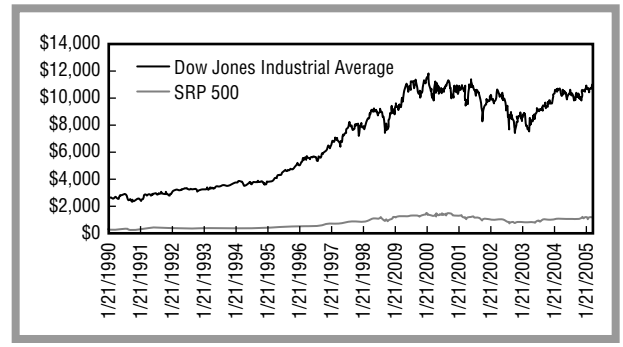


Figure 2.

X-Y SCATTER GRAPHS

At the fundamental level, all line charts are x-y scatter graphs. The x-y scatter graph is a plotting of all the coordinated points in a set of data. It does not usually have connecting lines. Consequently, the line chart is the next step after plotting the relationships between the x values and the y values. By simply connecting the dots, the XY scatter graph turns into a line chart.

X-y scatter graphs are useful when there are numerous variables to plot and a connecting line is not important. With the x-y scatter graph, clusters of data are easy to follow. These types of charts are used extensively in statistics to see the correlation of dependant and independent variables. As shown in Figure 4, a trend line could be drawn through the midpoint of all the data to show the aforementioned trending of different variables.

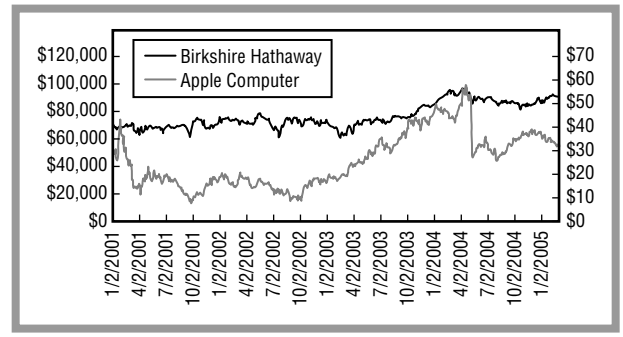


Figure 3.

COLUMN AND BAR CHARTS

Column charts are useful when sizing different categories of data and comparing them to each other. Column charts are usually used for fewer observations than seen in a line chart. For instance, to measure points scored per starting basketball player in a particular game, a column chart, such as Figure 5, would have a column (category) to represent each player. The players would be labeled along the x axis, and the points scored for each measured on the y axis.

To measure hours of sunlight per day over a one-year period, a column chart would need 365 x-axis categories. It is easier to show this with a line graph and have 12 x-axis categories, each measuring one month with the appropriate number of days between each category. Having 365 columns on one graph is usually unrealistic.

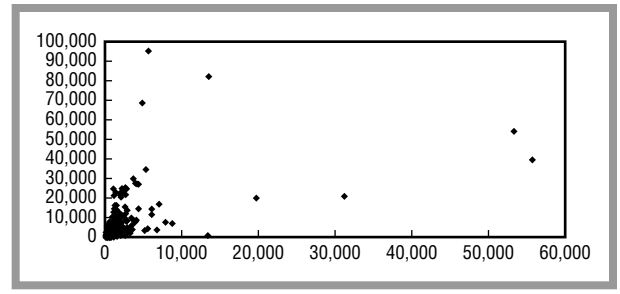


Figure 4.

BAR CHARTS

Bar charts are constructed in the same way as column charts with the difference being that the categories are plotted on the y axis and the data values are plotted on the x axis. This results in the columns protruding out horizontally, rather than up vertically. Because the column is now horizontal, it is referred to as a bar, like a parallel bar, rather than a column. Figure 6 represents a bar chart.

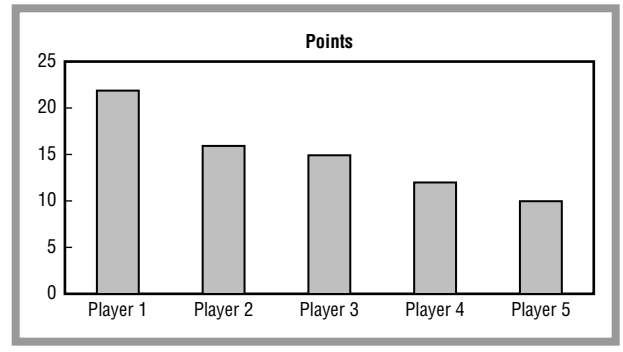


Figure 5.

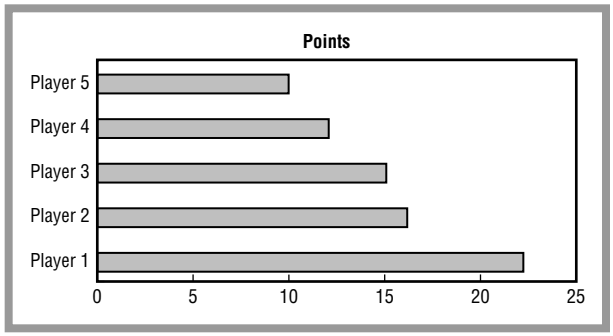


Figure 6.

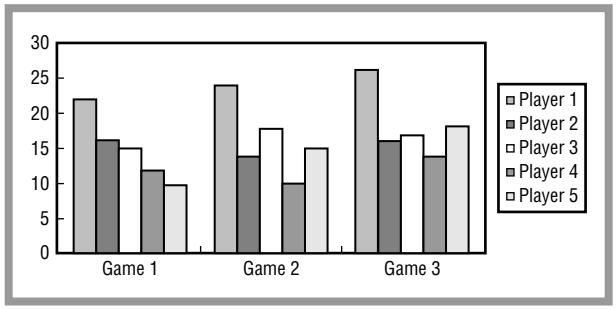


Figure 7.

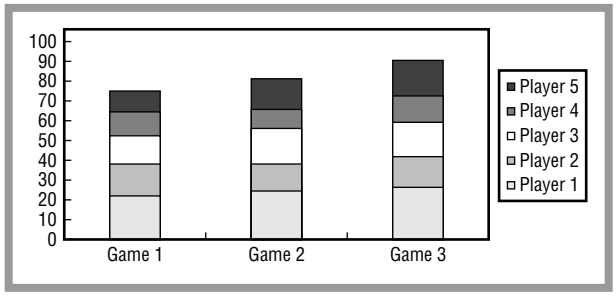


Figure 8.

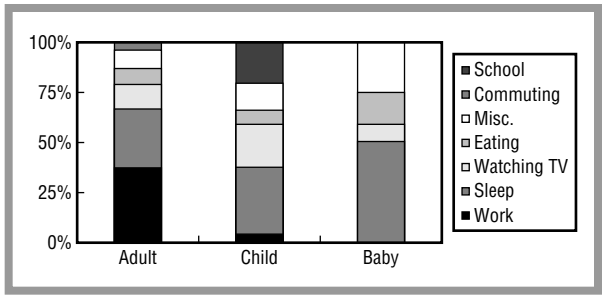


Figure 9.

CLUSTERED COLUMN CHARTS

Clustered columns are charts with the columns adjacent to each other, that is, they are touching, with multiple columns per category. For instance, when charting the points scored per player per game, each category would represent one game and there would be columns per category, all adjacent to each other. The next category would also have the same number of columns and would represent a different game. As Figure 7 shows, this type of chart needs a legend explaining what each column, or data series, represents.

Clustered column charts are very useful in comparing different categories as well as the series within the category. Often the individual columns are color coded as another way to separate and compare the data.

STACKED COLUMN CHARTS

The stacked column chart puts the different values per category on top of each other. This type of chart gives a relative sizing of each value per category. The stacked chart is good for comparing data both within categories and across categories. As in Figure 8, sometimes lines are drawn from the tops of each value between categories to help show the rate of change from one category to the next.

100% STACKED COLUMN CHARTS

Similar to the stacked chart, the 100% stacked chart gives a relative measurement categories, but sets the sum of all category values to 100%. The end result is a percentage measure of each value per category, rather than the actual value. The 100% stacked chart is useful when the overall category is a fixed amount. As an example, since there are 24 hours in the day, Figure 9 shows how those hours are spent on a percentage basis.

PIE CHARTS

Pie charts—much like the 100% stacked chart—convert all the data to percentages. That is, a pie chart sums all the different data categories and accepts this total to be 100% of the data, as in the whole pie. Then, each category is measured as a slice of the pie and represents therefore the percentage of the pie.

The formula for taking hypothetical data of 95 12th graders, 115 11th graders, 150 10th graders, and 180 9th graders is $95 + 115 + 150 + 180 = 540$. Therefore the pie calculations for 12th graders would be $95 / 540 = 18\%$. Figure 10 represents doing the same calculation for each grade, keeping the denominator at 540 and changing the numerator to represent the total students in each grade.

Pie charts are useful to help show how data sizes up to other data in one category. They also help to rank data quickly because they show which data category is largest, which is smallest, and all others in between.

A Brief History of Discovery and Development

Charts are sometimes referred to as Cartesian planes. This is the early name for a two-dimensional grid representation of numbers, first developed by French mathematician, philosopher, and scientist René Descartes (1596–1650).

Real-life Applications

People in finance and business often use charts, consequently they have develop elaborate charting techniques and different ways of displaying data in the three basic charting types.

In the field of finance there are people who make their living trading securities based solely on analyzing the charts of stocks, bonds, and commodities. This business is known as technical analysis, and the people who work in this field are known as chartists.

LINE CHARTS

Line charts are useful for charting large quantities of data, with numerous categories that would be impractical to view individually. For instance, to look at the daily closing price of a stock over a five-year period, it would require viewing approximately 1,260 individual data points (there are approximately 252 trading days in each calendar year). By using a line chart, each point can be graphed individually with tiny lines connecting the dots to give the visual quality of a continuous line. This will show trends over the five-year period and offers a sense of how the price has behaved during the entire time horizon or during specific time periods.

A more technical application for line charts is to plot a few data items and then calculate the slope of the line between each data point, or selective data points. This is useful in measuring volatility, and it also reveals trends and periods of drastic change more precisely.

COLUMN/BAR CHARTS

The term stack-up is pertinent in column charts because that is exactly what happens: data are stacked on top of each other or next to each other, for comparative purposes.

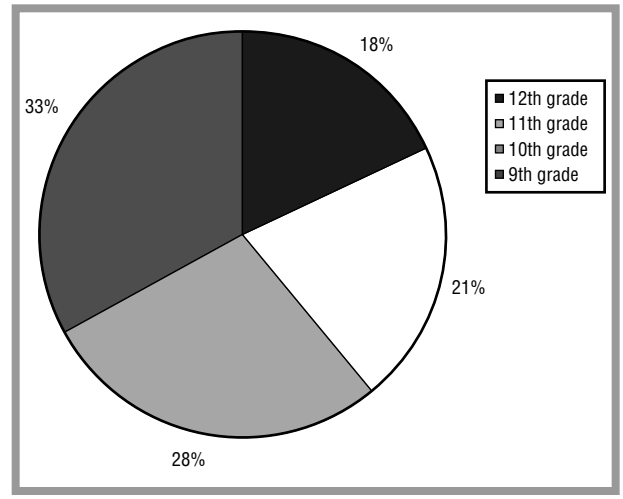


Figure 10.

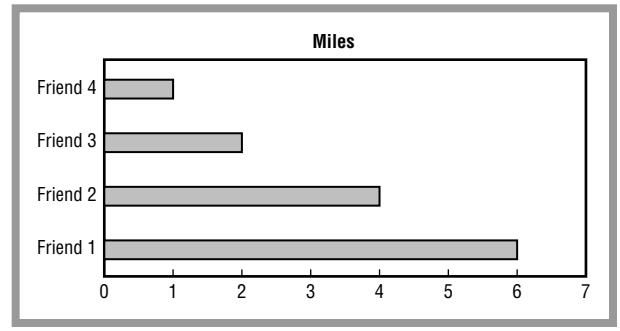


Figure 11.

Bar charts are often used to measure distance or growth. Since distance of travel is usually viewed from left to right, the length of each bar is a perfect way to visualize how far the data series goes. To compare distance traveled from one person's house to four other people's houses, Figure 11 is a bar chart that shows how far each trip is. It is easy to see which distance is furthest, which is shortest, and others in between.

PIE CHARTS

Pie charts are used to see how a particular data set is partitioned. A pie chart assumes that the data set is the whole universe of data, and it will show the individual percentages that make up that whole. Pie charts are most useful when the total is a fixed quantity.

When looking at total points scored in a football game, each player's points scored is shown as slices of the

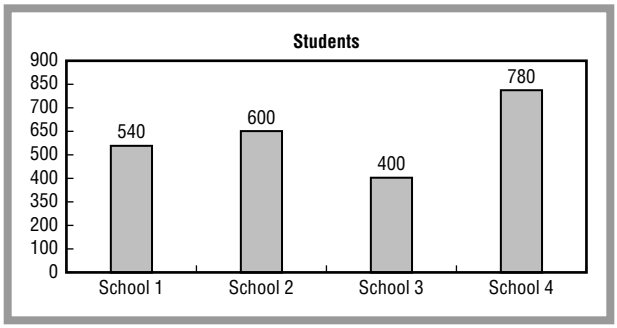


Figure 12.

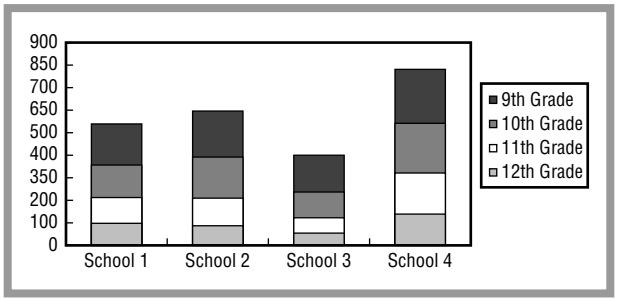


Figure 13.

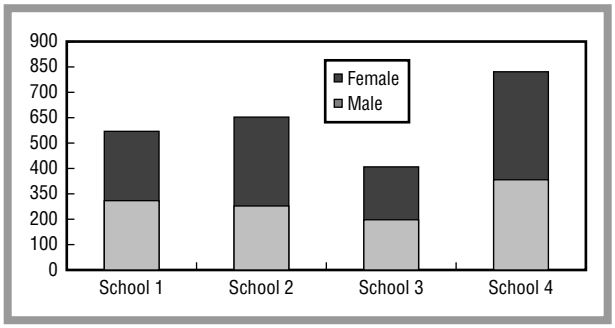


Figure 14.

pie, and the whole pie would represent the total points scored. Alternatively, to look at points scored by just three players, a pie chart is not useful, because other points could have been scored by different players, and the players do not represent the whole, they are only a fraction of the whole.

USING THE COMPUTER TO CREATE CHARTS

There are many computer programs that quickly do most chart plotting. The most common is Microsoft Excel, which has many different predetermined chart templates, based on the three basic charts, and formats data into a chart.

Excel and other charting programs have created preformatted charts to represent data in as many ways as possible, but at the root of all these charts are the three basic chart formats. One area where they have made significant changes in appearance is in area charts, or other three-dimensional chart types. While the basic charting procedure is basically the same, these charting programs have tried to add a third dimension, depth, to the basic two-dimensional chart. While this is helpful with very specific types of data, the two-dimensional charts are still the most commonly used.

CHOOSING THE RIGHT TYPE OF CHART FOR THE DATA

Organization of data is an important part of telling a story, and conveying that story to others. Charts are a quick way of showing the relational aspects of different categorized data sets; charts take the quantitative aspects of information and create a picture to make it easier for the viewer to quickly see relationships. Therefore, choosing the correct chart to represent data sets is a key element of conveying the story, and communicating how the data looks.

For example, at the beginning of the semester the math teacher makes the following announcement: the school administrators want to analyze the demographics of this high school relative to three other high schools in neighboring states. Furthermore, the administration has made the analysis a contest, and everyone in any math class is welcome to participate. All entries will be voted on fairly and independently. The teacher also states: if the winner is in a particular class, that participating student will receive an A for the course.

After collecting the data, the student ends up with the following information for all four schools: total students, broken out by grade; number of male and female students; total square feet of each school; number of teachers; number of classes offered; and the number of students who took the SAT tests, per state, over a 25-year period.

Using line, column, and pie charts, the data is organized in the following way: First, a basic column chart is created showing the total students for each school, as in Figure 12. Secondly, in Figure 13, a stacked bar chart is created, each with four columns, so each segment is representing one grade and each column is representing each school. Figure 14 represents this same concept used to show the distribution of males and females for each school.

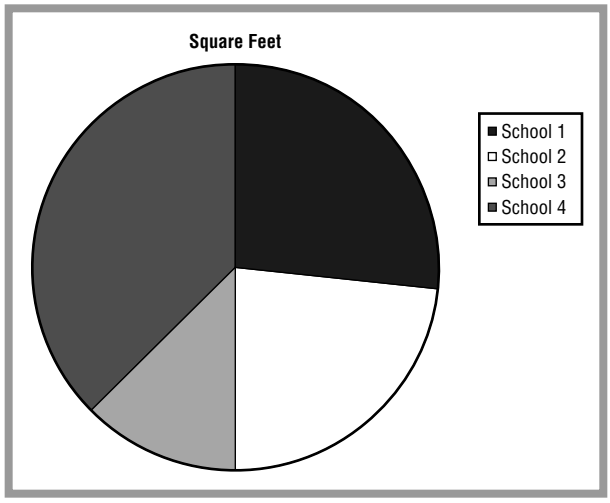


Figure 15.

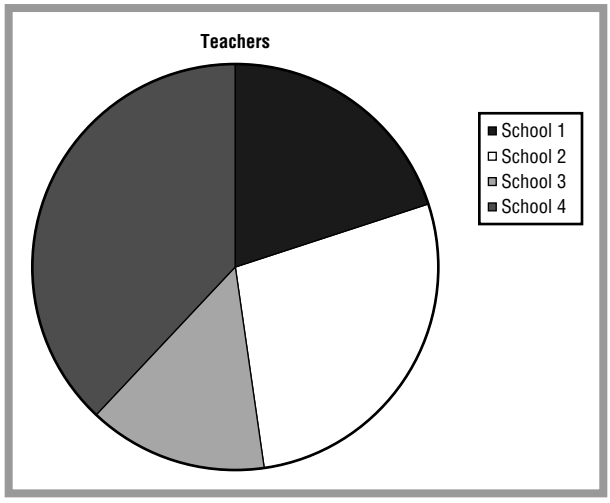


Figure 16.

Using a pie chart to plot the square feet per school, the pie chart has four segments, one for each school, and each segment of pie represents the percentage of square feet as a portion of the whole, as shown in Figure 15. Figure 16 represents a pie chart to plot the number of teachers for each school, and Figure 17 is the third pie chart that has the number of classes per school.

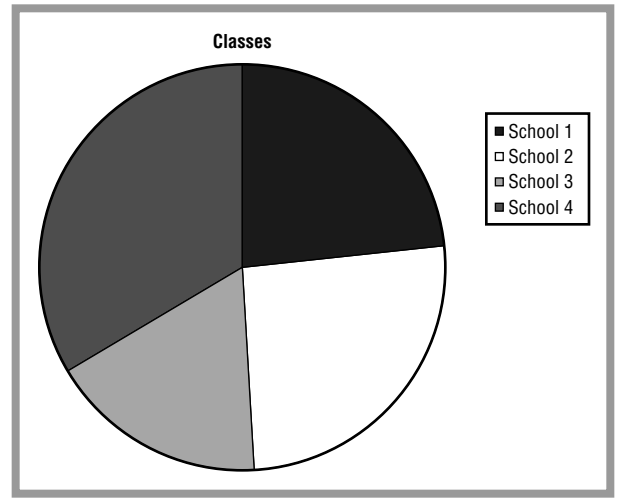


Figure 17.

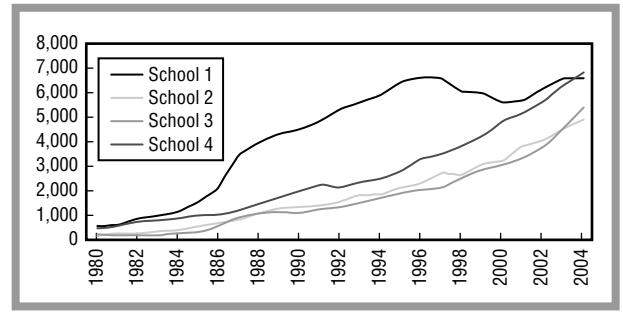


Figure 18.

Lastly, Figure 18 is a line chart used to plot the average SAT scores over the 25-year period. With 25 categories on the x axis, and the scores on the y axis, the data points are plotted, the dots connected, and a line chart is created that spans the 25-year period.

Where to Learn More

Books

Excel Charts. Somerset, NJ: John Wiley & Sons, 2005.

Key Terms

Dependant variable: What is being modeled; the output.

Independent variable: Data used to develop a model, the input.

Computers and Mathematics

Overview

Mathematics is integral to computers. Most computer processes and functions rely on mathematical principles. The word "computers" is derived from computing, meaning the process of solving a problem mathematically. Large complex calculations (or computing) in engineering and scientific research often require basic calculators and computers.

Computers have evolved greatly over the years. These days, computers are used for practically anything under the Sun, education, communication, business, shopping, or entertainment. Mathematics forms the basis of all these applications.

Applications of mathematical concepts are seen in the way computers process data (or information) in the form of bits, bytes, and codes, store large quantities of data by compression, and send data from one computer to another by transmission. With the advent of the Internet, communication has become extremely easy. Every computer is assigned a unique identity, using mathematical principles, making communication possible. In addition, mathematics has also found other applications in computers, such as security and encryption.

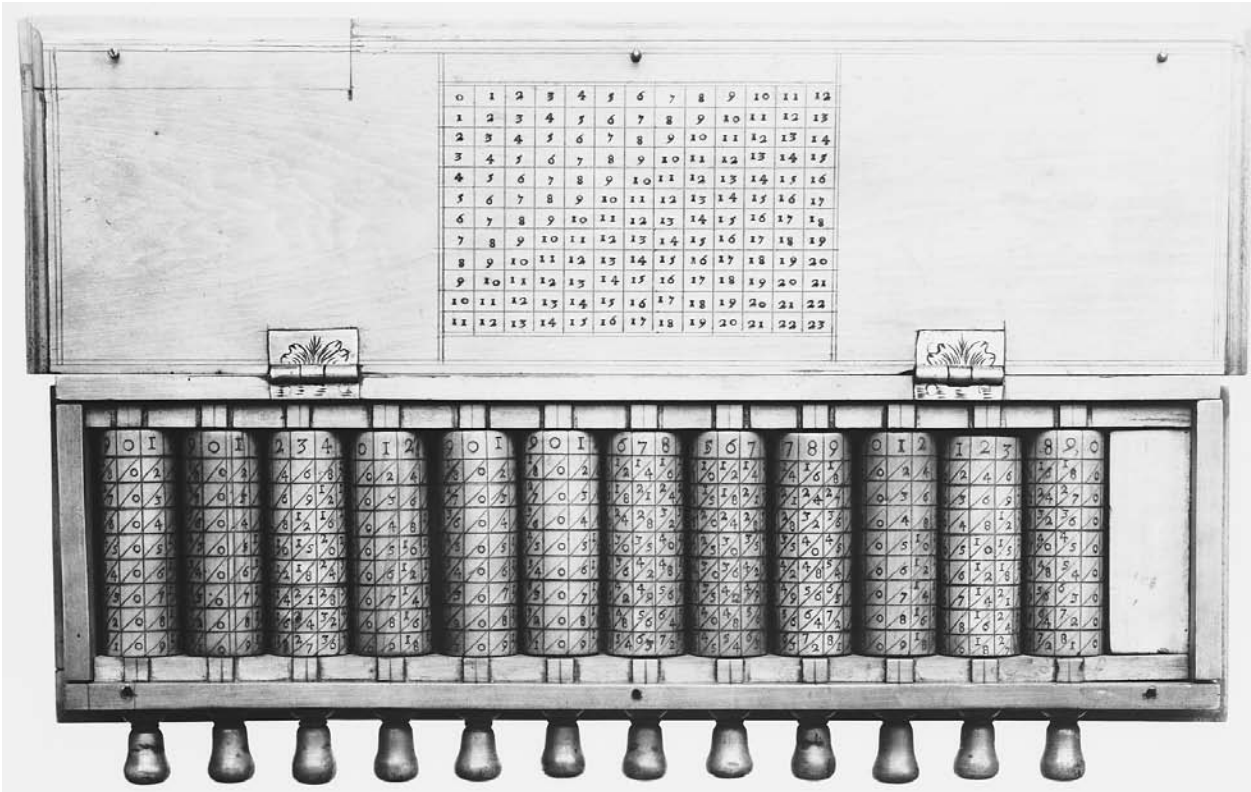
Fundamental Mathematical Concepts and Terms

BINARY SYSTEM

All computers or computing devices think and process in binary code, a binary number system. In a binary number system, everything is described using two values—on or off, true or false, yes or no, one or zero, and so on. The simplest example of a binary system is a light switch, which is always either on or off. A computer contains millions of similar switches. The status of each switch in the computer represents a bit or binary digit. In other words, each switch is either on or off. The computer describes one as "on" and zero as "off."

Any number can be represented in the binary system as a combination of zeros and ones. In the binary number system, each number holds the value of increasing powers of two, e.g., 2^0, 2^1, and so on. This makes counting in binary easy. The binary representation for the numbers one to ten can be shown as follows:

- 0 = 0
- 1 = 1
- 2 = 10
- 3 = 11



A calculating device created by Scottish mathematician John Napier in 1617 which consists of cylinders inscribed with multiplication tables. It's also known as "Napier's Bones." BETTMANN/CORBIS.

- 4 = 100
- 5 = 101
- 6 = 110
- 7 = 111
- 8 = 1000
- 9 = 1001
- 10 = 1010.

ALGORITHMS

The key principle in all computing devices is a systematic process for completing a task. In mathematics, this systematic process is called an algorithm. Algorithms are common in daily life as well. For example, when building a house, the first step involves building the floor base (or foundation), followed by the walls, and then the ceiling or roof. This systematic procedure to solve the problem of building a house is an example of an algorithm.

In a nutshell, algorithms are a list of step-by-step instructions. In mathematical terms, these are also sometimes known as theorems. A computer program, or application, is made up of a number of such algorithms. Besides, every process in a computer also depends on a specific algorithm. For example, when switching on the computer, the computer does what is known as "booting." Booting helps in properly loading the operating system (Windows, Mac, Dos, UNIX, and so on). During booting, the computer follows a set of instructions (defined by an algorithm). Similarly, while opening any program (say, MS Word), the computer is again instructed to follow a set of tasks so that the program opens properly.

Like complex mathematical problems, even the most complex software programs are based on numerous algorithms.

A Brief History of Discovery and Development

Although the modern computer was built only in the twentieth century, many primitive forms of the computer were used in ancient times. The early calculators can also be considered as extremely basic computers based on similar mathematical concepts. The word calculator, is derived from the Latin word *calculus* (or a small stone). Early

human civilizations used small stones for counting. Counting boards made up of stones were used for basic arithmetic tasks such as addition, subtraction, and multiplication.

This led to development of devices that enabled calculation of more complex numbers, and in quick time. With the progress of civilization, man saw the development of the abacus, the adding machine, the Babbage, and the prototype mainframe computers.

Modern computers, however, were invented in the twentieth century. In 1948, the mathematician Claude Shannon (1916–2001), working at Bell Laboratories in the United States, developed computing concepts that would form the basis of modern information theory. Shannon is often known as the father of information science. Computers were earlier only used by government institutions. Home or personal computers (known as PCs) came much later in the late 1970s and 1980s.

Today, personal computers and servers with a microprocessor chip (a small piece of computer hardware) are embedded in almost all lifestyle electronic products, from the washing machine and television to calculators and automobiles. Many of these chips are capable of computing in the same capacity as some basic computers. The advancement of mathematical concepts and theories has made it possible to develop sophisticated computers in smaller and smaller sizes, such as those found in handheld computers like the PDA (personal data assistant) and PMP (personal media player).

Ciphers, codes, and secret writing based on mathematical concepts have been around since ancient times. In ancient Rome, they were used to communicate secrets over long distances. Such codes are now used extensively in the field of computer science.

Real-life Applications

BITS

The bit is the smallest unit of information in a computer. As discussed earlier, a bit is a basic unit in a binary number system. A bit or binary digit stands for true or false, one or zero, on or off. The computer is made up of numerous switches. Each switch has two states (on and off). The value of each state represents a bit.

Bits are the basic unit of storage in computers. In other words, all data is stored in the form of bits. The reason for using a binary number system rather than decimal system for storage (and other purposes) is that with prevailing technology, it is much easier to implement the binary system in computers. Implementing the binary system is significantly cheaper, as well.

The speed of the computer (processor speed) in terms of processing applications is related to many factors, including memory space (also known as random access memory, or RAM). Most home computers are either 32-bit or 64-bit; 32-bit and 64-bit are the sizes of the memory space.

BYTES

In computers, bits are bundled together into manageable collections called bytes. A byte consists of eight bits. Bits and bytes are always clubbed together like atoms and molecules. Computers are designed to store data and process instructions in bytes. To handle large quantities of information (or bits), other units such as kilobytes, megabytes, and gigabytes are used. One kilobyte (KB) = 1,024 bytes = 2^{10} bytes (and not 1,000 bytes as commonly thought). Similarly, 1 megabyte (MB) = 1,048,576 bytes = 2^{20} bytes, and 1 gigabyte (GB) = 1,073,741,824 bytes = 2^{30} bytes.

The first computers were 1-byte machines. In other words, they used octets or 8-bit bytes to store information, and they represented 256 values (2^8 values, integers zero to 255).

The latest computing machines are 64-bit (or eight bytes). This type of representation makes computing easier in terms of both storage and speed. Bits and bytes form the basis of many other computer processes and functions. These include CD storage, screen resolution, text coding, data comparison, data transmission, and much more.

TEXT CODE

All information in the computer is stored in the form of binary numbers. This includes text, as well. In other words, text is not stored as text, but as binary numbers. The rule that governs this representation is known as ASCII (American Standard Code for Information Interchange). The ASCII system assigns a code to every letter of the alphabet (and other characters). This code is stored as a seven digit binary number in computers. Moreover, the ASCII code for a capital letter is different than the code for the small letter. For example, the ASCII code for "A" is 10, whereas that for "a" is 97. Consequently, the value of "A" is stored as 0001010 (its binary representation), whereas "a" is 1100001.

Every character is stored as eight bits (a leading bit in addition to the seven bits for the ASCII code), or one byte. Thus, the word "happy" would require five bytes. An entire page with 20 lines and 60 characters per line would require 1,200 bytes.

The main benefit of storing text code as binary numbers is that it makes it easier for the computer to store and process the data. Besides, mathematical operations can be performed on binary representations of text.

PIXELS, SCREEN SIZE, AND RESOLUTION

A pixel is derived from the words picture and element. The smallest and the most basic unit of images in computers is the pixel. A pixel is a tiny square block. Images are made up of numerous pixels. The total number of pixels in a computer image is known as the resolution of the image. For example, a standard computer monitor displays images with the resolution 800 × 600. This simply means that the image (or the entire computer screen) is 800 pixels wide and 600 pixels high.

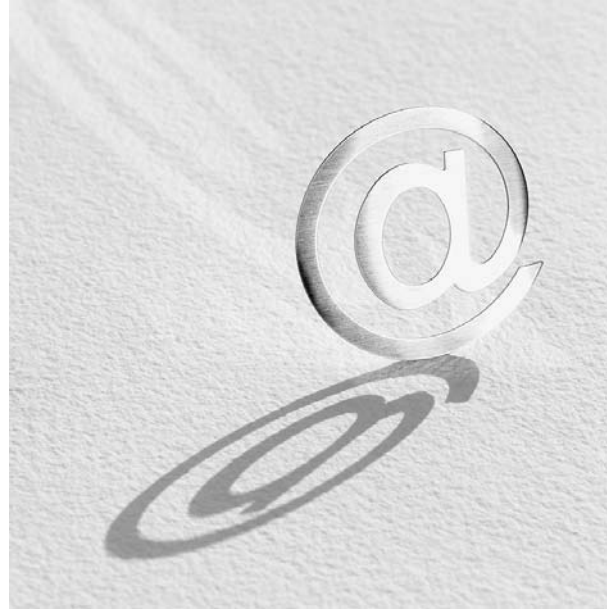
Each pixel is also stored as eight bits (or one byte). Again, its representation is in the form of binary numbers. Storing the value of the color of a pixel is far easier in binary format, as compared with other formats. The maximum number of combinations of zeros and ones in an 8-bit number is 256 (2^8). Each combination represents a color. Simply put, every pixel can have one of 256 different colors.

This kind of computer display is called an "8-bit" or "256-color" display, and was very common in computers built in the 1990s. In contrast, newer computer monitors built after the year 2000 have a significantly higher number of colors (in millions). These are the 16-bit and 24-bit monitors.

The color of every pixel in a computer image is a combination of three different colors—red, green, and blue (RGB). RGB is common terminology used in computer graphics and images, and simply means that every color is a combination of some portion of red, green, and blue colors. The value of each of these colors is stored in one byte. For example, the color of a pixel could be 100 of red, 155 of green, and 200 of blue. Each of these values is stored in binary format in a byte. Note that the color values can range from zero to 255. Thus, every color pixel has three bytes. Subsequently, a computer monitor with the resolution 800 × 600 would need 3 × 800 × 600, or 1,440,000 bytes.

IP ADDRESS

Every computer on a network has a specific address. A number, known as the Internet protocol address, or IP address, indicates this. The reason for having an IP address is simple. To send a packet or a letter through regular mail, the address of the recipient is required. Similarly, for communicating with a computer (from another computer), the address of that computer is required. Every computer has a unique IP address that clearly distinguishes it from other computers. The concept of the IP address is based on mathematical principles, and there are rules that govern the value of the IP address. For example, an IP address is always a set of four numbers separated by dots (e.g., 204.65.130.40).



Internet mathematics translates binary code into web addresses and other information. ROYALTY-FREE/CORBIS.

Remember, the computer only understands binary numbers. Consequently, the IP address is also represented as a binary number. The binary representation is octet (equivalent to the representation of a byte). Technically, every IP address is a 32-bit number divided into four bytes, or octets (eight bites). Each octet represents a specific number. For example, in the above case, 204 would be stored in one octet, 65 in another octet, and so on. The binary representation (as stored in the computer) for the above-mentioned IP address would be: 11001100 .01000001.10000010.0101000.

Communication between computers becomes far easier with binary representation. The IP address consists of two components, the network address and the host address. The network address (the first two numbers) represents the address of the entire network. For example, if a computer is part of a network of computers connected into an entire company, the first two numbers would represent the IP address of the company. In other words, for all computers connected to the company network, the first two numbers would remain the same.

The host address (the last two numbers) represents the address of a computer specifically. For example, the third number might represent a particular department within a company, whereas the last number would represent a particular computer in that department. Consequently, two computers within the same department (and part of the same company) would have the same first three numbers. Only the last number would be different. Similarly, two computers that are part of different departments would have the same first two numbers.

As each number in the IP address is allowed a maximum of one octet (or eight bites), the maximum value the number can have is 255. In other words, the values of every number in the IP address ranges from zero to 255. An IP address that contains a number higher than this range would be incorrect. For example, 204.256.12.0 is incorrect, as 256 is not valid.

SUBNET MASK

With the advent of the Internet, the number of computers that are connected worldwide is quickly rising. The Internet is a huge network of computers. Subsequently, each computer has an IP address that helps it communicate with the rest. For example, to send an email, the email address must be entered. This email address is translated to a specific IP address, that of the recipient. As of 2005, there are millions of computers connected to the Internet. As mentioned earlier, IP addresses have a limitation. Each number can only have a value within a specific range (zero to 255).

The IP address given to any computer on the Internet is temporary. In other words, as soon as a computer connects to the Internet, it receives a unique IP address. As soon as the Internet is disconnected, this IP address is free and can be used by another computer. When the same computer connects again, it would get another IP address. With the high number of computers connected to the Internet simultaneously, it is difficult to accommodate every computer within this range. This is where the concept of Subnet mask comes in.

Subnets, as the name suggests, are sub-networks. The host address (from the IP address) is divided into further subnets to accommodate more computers. This is done in such a way that a part of the host address identifies the subnet. The subnet is also shown as a binary number. Communication becomes easier because of the binary representation.

Take, for example, the IP address 204.65.130.40. Its binary equivalent is 11001100.01000001.10000010 .00101000.

The subnets would have the same network address (first two numbers). The first four bits of the host address (third number) would be the same as well, to identify the host of the subnet. In this case, 1000 would be unchanged. The remaining four bits of the host address would be unique to each subnet. Every subnet, in turn, can have numerous computers. Every computer on the subnet would have a unique fourth number in the IP address. Consider the following scenario:

The main IP address is 11001100.01000001 .10000010.00101000. This could have many subnets such as 11001100.01000001.10000111.00111010, 11001100 .01000001.10000101.0100010, and so on. Note that the first four digits of the third number (host address) are same but the remaining are different, indicating different subnets on the same host. The fourth number indicates a specific computer on the subnet. For computers on the same subnet, the first three numbers would remain the same.

Simply put, the subnet mask ensures that more computers can be accommodated within a network. Every subnet mask number identifies the network address, the host, the subnet, as well as the computer.

COMPRESSION

Computers store (and process) data that include numbers, arithmetic calculations, and words. In addition, the data may also be in the form of pictures, graphics, and videos. In computers, data is stored in files. File sizes, depending on the type of data, can be huge. Many times the size of a file becomes unmanageable. In such cases, better ways of storing and process data, must be used. Given below are some comparisons to provide a better understanding of sizes of different files on a computer.

One alphabetic character is represented by one byte, one word is equivalent to eight to ten bytes or so, a page averages about two kilobytes, an entire book averages one megabyte or more, twenty seconds of good quality video occupy anywhere from two to ten megabytes, and so on. Similarly, a compact disc (CD) has 600–800 megabytes of data.

Storing such huge amounts of information in a computer can often be difficult. Besides, it is almost impossible to send large data from one computer to another through e-mail or other similar means. Moreover, downloading a significant amount of data from the Internet (such as movie files, databases, application programs) can be extremely time consuming, especially if using a slow dial up connection. This is where compression of the data into a manageable size becomes important.

Certain applications based on mathematical algorithms compress the data. This allows the basic data that a computer sees in binary format, to be stored in a compressed format requiring much lower storage space. Compressed data can be uncompressed using the same application and algorithm.

Compression is extremely beneficial, especially when a large file has to be sent from one computer to another. In case of e-mail, sending a one-megabyte (MB) file through a dial up connection, would take considerable time, anywhere from fifteen to thirty minutes. Bigger files would take even longer. Besides, e-mails might not have the capacity of sending (or receiving) bigger files. In such cases, sending zipped files that are much smaller is useful. Similarly, downloading compressed files from the Internet rather than the large original ones is a better option.

There are also other types and methods for compressing. Run length compression is another type that is used widely. In run length compression, large chunks, or runs, of consecutive identical data values are taken, and each of these is replaced by a common code. In addition to the code, the data value and the total length are also recorded. Run length compression can be quite effective. However, it is not used for certain types of data such as text, and executable programs. For these types of files, run length compression does not work. Without going into the technical specifics of run length compression, this method works quite well on certain types of data (especially images and graphics), and is subsequently applied to many data compression algorithms. Most compressed files can be uncompressed to obtain the original. However, in almost all cases, some data is lost in the process. For visual and audio data, some loss of quality is allowed without losing the main data. By taking advantage of limitations of the human sensory system, a great deal of space is saved while creating a copy that is very similar to the original. In other words, although compression results in some data loss, this loss can be insignificant and the naked eye usually cannot usually discern the difference between the original and the un-compressed file. The defining characteristics of these compression methods are their compression speed, the compressed size, and the loss of data during compression.

Apart from computers, compression of images and video is also used in digital cameras and camcorders. The main purpose is to reduce the size of the image (or video) without compromising on the quality. Similarly, DVDs also use compression techniques based on mathematical algorithms to store video.

In audio compression, compression methods remove non-audible (or less audible) components of the signal while compressing. Compression of human speech is sometimes done using algorithms and tools that are far more complex. Audio compression has applications in Internet telephony (voice chat through the internet), audio CDs, MP3 CDs, and more.

DATA TRANSMISSION

In computing, data transmission means sending a stream of data (in bits or bytes) from one location to another, using different technologies. Two of these technologies are coding theory and hamming codes. These are both based on algorithms and other mathematical concepts.

Coding theory ensures data integrity during transmission. In other words, it ascertains that the original data is safely received, without any loss. Messages are usually not transmitted in their original form. They are transmitted in coded or encrypted form (described later). Coding theory is about making transmitted messages easy to read. Coding theory is based on algorithms. In 1948, the mathematician Claude Shannon presented coding theory by showing that it was possible to encode in an effective manner. In its simplest form, a coded message is in the form of binary digits or bits, strings of zero or one. The bits are transmitted along a channel (such as a telephone line). While transmitting, a few errors may occur. To compensate for the errors, more bits of information than required are generally transmitted.

The simplest method (part of the coding theory developed by Shannon) for detecting errors in binary data is the parity code. Concisely, this method transmits an extra bit, known as the parity bit, after every seven bits from the source message. However, the parity code method can merely detect errors, not correct them. The only method for correcting them is to ask for the data to be transmitted again.

Shannon developed another algorithm, known as the repetition algorithm, to ensure detection as well as correction of errors. This is accomplished by repeating each bit a specific number of times. The recipient sees which value (zero or one) occurred more often and assumed that was the actual value. This process can detect and correct any number of errors, depending on how many repeats of each bit are sent. The disadvantage of the repetition algorithm is that it transmits a high number of bits, resulting in a considerable amount of repetitive bits. Besides, the assumption that a bit that is received more often, is the actual bit, may not hold true in all cases.

Another mathematician, Richard Hamming (1915–1998), built more complex algorithms for error correction. Known as Hamming codes, these were more efficient, even

with very low repetition. Initially, Hamming produced a code (based on an algorithm) in which four data bits were followed by three check bits that allowed the detection and the correction of a single error. Although, the number of additional bits is still high, it is without a doubt lower than the total number of bits transmitted by the repetition algorithm. Subsequently, these additional bits (check bits) were reduced even further by improving the underlying algorithms. Hamming codes are commonly used for transmitting not just basic data (in the form of simple email messages), but also more complex information.

One such example is astronomy. The National Aeronautics and Space Administration (NASA) uses these techniques while transmitting data from their spacecrafts back to Earth (and vice versa). Take, for example, the NASA *Mariner* spacecraft sent to Mars in the 1960s. In this case, coding and error correction in data transmission was vital, as the data was sent from a weak transmitter over very long distances. Here the data was read perfectly using the Hamming code algorithm. In the late 1960s and early 1970s, the NASA *Mariner* sent data using more advanced versions of the Hamming and coding theories, capable of correcting seven errors out of thirty-two bits transmitted. Using this algorithm, over 16,000 bits per second of data was successfully relayed back to Earth.

Similar data transmission algorithms are used extensively for communication through the Internet since the late 1990s. The Hamming codes are also used in preparing compact discs (CDs). To guard against scratches, cracks, and similar damage, two overlapped Hamming codes are used. These have a high rate of error correction.

ENCRYPTION

Considerable confidential data is stored and transmitted from computers. Security of such data is essential. This can be achieved through specialized techniques known as encryption. Encryption converts the original message into coded form that cannot be interpreted unless it is de-coded back to the original (decryption). Encryption, a concept of cryptography, is the most effective way to achieve data security. It is based on complex mathematical algorithms.

Consider the message abcdef1234ghij56789. There are several ways of coding (or encrypting) this information. One of the simplest ways is to replace each alphabet by a corresponding number, and vice versa. For example, “a” would become “1”, “b” would be “2”, and so on. The above original message can, thus be encrypted as 123456abcd78910 efghi. The message is decrypted using the same process and converted back in the original form.

Complex mathematical algorithms are designed to create far more complex encryption methods. The information regarding the encryption method is known as the key.

Cryptography provides three types of security for data:

- Confidentiality through encryption—This is the process mentioned above. All confidential data is encrypted using certain mathematical algorithms. A key is required to decrypt the data back into its original form. Only the right people have access to the key.
- Authentication—A user trying to access coded or protected data must authenticate himself/herself. This is done through his/her personal information. Password protection is a type of authentication that is widely used in computers and on the Internet.
- Integrity—This type of security does not limit access to confidential information, as in the above cases. However, it detects when such confidential is modified. Cryptographic techniques, in this case, do not show how the information has been modified, just that it has been modified.

There are two main types of encryption used in computers (and the Internet)—asymmetric encryption (or public-key encryption), and symmetric encryption (or secret key encryption). Each of these is based on different mathematical algorithms that vary in function and complexity.

In brief, public key encryption uses a pair of keys, the public key, and the private key. These keys are complimentary, in the sense that a message encrypted using a particular public key can only be decrypted using a corresponding private key. The public key is available to all (it is public). However, the private key is accessible only by the receiver of a data transmission. The sender encrypts the message using the public key (corresponding to the private key of the receiver). Once the receiver gets the data, it is decrypted using the private key. The private key is not shared with anyone other than the receiver, or the security of the data is compromised.

Alternatively, symmetric secret key encryption relies on the same key for both encryption and decryption. The main concern in this case is the security of the key. Subsequently, the key has to be such that even if someone gets hold of it, the decryption method does not become too obvious. For this purpose, encryption and decryption algorithms for secret key encryption are quite complex.

The key, as expected, is shared only by the receiver and the sender (unlike public key encryption, where everyone knows the public key). The key can be anything ranging from a number, a word, or a string of jumbled up letters and other characters. In simple terms, the original

Key Terms

Bit: The smallest unit of storage in computers. A bit stores a binary value.

Byte: A byte is a group of eight bits.

Encryption: Using a mathematical algorithm to code a message or make it unintelligible.

Pixel: Short for "picture," a pixel is the smallest unit of a computer graphic or image. It is also represented as a binary number.

data is encoded using a simple or complex technique defined by a mathematical algorithm. The key also holds the information on how the algorithm works. The same algorithm can then be used to decode the message back into its original form.

Encryption is used frequently in computers. Most data is protected using one of the above mentioned encryption techniques. The Internet also widely applies encryption. Most websites protect their content using these methods. In addition, payment processing on websites also follows complex encryption algorithms (or standards) to protect transactions.

Where to Learn More

Books

Cook, Nigel P. *Introductory Computer Mathematics*. Upper Saddle River, NJ: Prentice Hall, 2002.

Graham, Ronald H., et al. *Concrete Mathematics: A Foundation for Computer Science*. Boston, MA: Addison-Wesley, 1994.

Conversions

Overview

Conversion is the process of changing units of measurement from one system to another. The ability to convert units such as distance, weight, and currency is an increasingly important skill in an emerging global economy. In area of research and technological applications such as science and engineering, the ability to convert data is crucial.

No better example of how critical a role conversion math can play can be found in the destruction of NASA's *Mars Climate Orbiter* in 1999. The *Mars Climate Orbiter* was one of a series of NASA missions in a long-term program of Mars exploration known as the Mars Surveyor Program. The orbiter mission was designed to have the orbiter fire its main engine to enter into orbit around Mars at an altitude of about 90 miles (about 140 km). However, a series of errors caused the probe to come too close to Mars and, as a result, the probe was only about 35 miles (57 km) from the Martian surface when it attempted to enter orbit—an altitude far below the minimum safe altitude for orbit. As a result the *Mars Climate Orbiter* is presumed to have been destroyed as it reentered the Martian atmosphere.

Engineering teams contracted by NASA used different measurement systems (English and metric) and never converted the two measurements. As a result, the probe's attitude adjustment thrusters failed to fire properly and the probe drifted off course toward its fatal demise.

Fundamental Mathematical Concepts and Terms

In addition to traditional English measurements, International System of Units (SI) and MKS (meter-kilogram-second) units are part of the metric system, a system based on powers of ten. The metric system is used throughout the world—and in most cases provides the standard for measurements used by scientists. On an everyday basis, nearly everyone is required to convert values from one unit to another (e.g., the conversion from kilometers per hour to miles per hour).

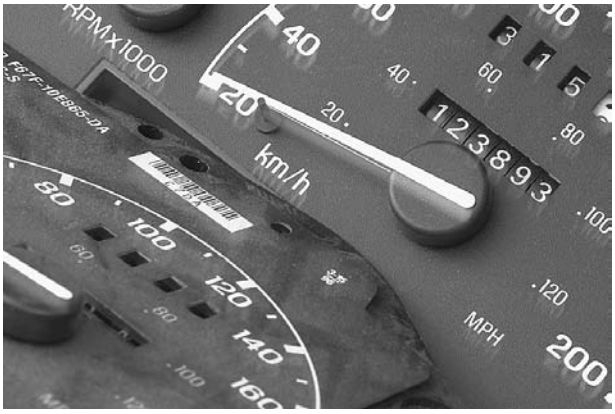
This need for conversation applies widely across society, from fundamental measurement of the gap in spark plugs to debate and analysis over sports records.

When values are multiplied or divided, they can each have different units. When adding or subtracting values, however, the values must added or subtracted must have the same units. A notation such as "ms^{-1}" is simply a different way of indicating m/s (meters per second).

Units must properly cancel to yield a proper conversion. If an Olympic sprinter runs 200-meter race in 19.32 seconds, he runs at an average speed of average speed of 10.35 meters per second [200 m / 19.32 s = 10.35 m/s]. If a student wishes to convert this to miles per hour the conversion should be carried out as follows: (10.35 m/s) (1 mile / 1,609 m) (3,600 s / 1 hr) = 23.2 miles/hr. The units cancel as follows: (10.35 $^{m}/^{s}$) (1 mile / 1,609 ^{m}) (3,600 ^{s} / 1 hr) = 23.2 miles/hr.

Students should remember to be cautious when dealing with units that are squared, cubed, or that carry another exponent. For example, a cube that is 10 cm on each side has a volume that is expressed as a cube value (e.g., m^3 that is determined from multiplying the cube's length times the width times the height: V = (10 cm) (10 cm)(10 cm) = 1,000 cm^3.

Many conversions are autoprogrammed into calculators—or are easily made with the use of tables and charts.



Odometers sit in a shop that legally converts odometers from kilometers to miles in used cars imported from Canada. AP/WIDE WORLD PHOTOS. REPRODUCED BY PERMISSION.

THE METRIC UNITS

The SI starts by defining seven basic units: one each for length, mass, time, electric current, temperature, amount of substance, and luminous intensity. ("Amount of substance" refers to the number of elementary particles in a sample of matter. Luminous intensity has to do with the brightness of a light source.) However, only four of these seven basic quantities are in everyday use by non-scientists: length, mass, time, and temperature.

The defined SI units for these everyday units are the meter for length, the kilogram for mass, the second for time, and the degree Celsius for temperature. (The other three basic units are the ampere for electric current, the mole for amount of substance, and the candela for luminous intensity.) Almost all other units can be derived from the basic seven. For example, area is a product of two lengths: meters squared, or square meters. Velocity or speed is a combination of a length and a time: kilometers per hour.

Because the meter (1.0936 yd) is much too big for measuring an atom and much too small for measuring the distance between two cities, we need a variety of smaller and larger units of length. But instead of inventing different-sized units with completely different names, as the English-American system does, metric adaptations are accomplished by attaching a prefix to the name of the unit. For example, since kilo- is a Greek form meaning a thousand, a kilometer is a thousand meters. Similarly, a kilogram is a thousand grams; a gigagram is a billion grams or 10^9 grams; and a nanosecond is one billionth of a second or 10^{-9} second.

THE ENGLISH SYSTEM

In contrast to the metric system's simplicity stands the English system of measurement (a name retained to honor the origin of the system) that is based on a variety of standards (most completely arbitrary).

There many English units, including buckets, butts, chains, cords, drams, ells, fathoms, firkins, gills, grains, hands, knots, leagues, three different kinds of miles, four kinds of ounces, and five kinds of tons. There are literally hundreds more. For measuring volume or bulk alone, the English system uses ounces, pints, quarts, gallons, barrels and bushels, among many others.

THE INTERNATIONAL SYSTEM OF UNITS (SI)

The metric system is actually part of a more comprehensive International System of Units, a comprehensive set of measuring units. In 1938, the 9th General [International] Conference on Weights and Measures, adopted the International System of Units. In 1960, the 11th General Conference on Weights and Measures modified the system and adopted the French name *Système International d'Unités*, abbreviated as SI.

Nine fundamental units make up the SI system. These are the meter (abbreviated m) for length, the kilogram (kg) for mass, the second (s) for time, the ampere (A) for electric current, the Kelvin (K) for temperature, the candela (cd) for light intensity, the mole (mol) for quantity of a substance, the radian (rad) for plane angles, and the steradian (sr) for solid angles.

DERIVED UNITS

Many physical phenomena are measured in units that are derived from SI units. As an example, frequency is measured in a unit known as the hertz (Hz). The hertz is the number of vibrations made by a wave in a second. It can be expressed in terms of the basic SI unit as s^{-1}. Hertz units are used to describe, measure, and calibrate radio wavelengths and computer processing speeds.

Pressure is another derived unit. Pressure is defined as the force per unit area. In the metric system, the unit of pressure is the Pascal (Pa) and can be expressed as kilograms per meter per second squared, or kg/m s^2. Measurements of pressure are important in determining whether gaskets and seals are properly placed on automobile motors or properly functioning in air-conditioning units.

Even units that appear to have little or no relationship to the nine fundamental units can, nonetheless, be expressed in terms of those units. The absorbed dose, for example, indicates that amount of radiation received by a person or object. In the metric system, the unit for this measurement is the "gray." One gray can be defined in terms of the fundamental units as meters squared per second squared, or m^2 / s^2.

Many other commonly used units can also be expressed in terms of the nine fundamental units. Some of the most familiar are the units for area (square meter: m^2), volume (cubic meter: m^3), velocity (meters per second: m/s), concentration (moles per cubic meter: mol/m^3), and density (kilograms per cubic meter: kg/m^3).

As previously mentioned, a set of prefixes is available that makes it possible to use the fundamental SI units to express larger or smaller amounts of the same quantity. Among the most commonly used prefixes are milli- (m) for one-thousandth, centi- (c) for one-hundredth, micro- (μ) for one-millionth, kilo- (k) for one thousand times, and mega- (M) for one million times. Thus, any volume can be expressed by using some combination of the fundamental unit (liter) and the appropriate prefix. One million liters, using this system, would be a megaliter (ML) and one millionth of a liter, a microliter (μL).

UNITS BASED ON PHYSICAL OR "NATURAL" PHENOMENA

In the field of electricity the charge carried by a single electron is known as the elementary charge (e) and has the value of $1.6021892 \times 10^{-19}$ coulomb. This is termed a "natural" unit.

Other real-world or "natural" units of measurement include the speed of light (c: 2.99792458×10^8 m/s), the Planck constant (6.626176×10^{-34} joule per hertz), the mass of an electron (m_e: $0.9109534 \times 10^{-30}$ kg), and the mass of a proton (m_p: $1.6726485 \times 10^{-27}$ kg).

Each of the above units can be expressed in terms of SI units, but they are often also used as basic units in specialized fields of science.

A Brief History of Discovery and Development

Because the United States is the world's leading producer in many items, regardless of the near universal acceptance of the SI, the most frequent conversions between units are between the English system of weights and measures to those of the metric system. The metric system of measurement, first advanced and adopted by the France in the late eighteenth and early nineteenth century, has grown to become the internationally agreed-upon set of units for commerce, science, and engineering.

The United States is the only major economic power to yet fully embrace the metric system. The history of the metric system in the United States is bumpy, with progress toward inevitable metrification coming slowly over two centuries.

As early as 1800, U.S. government agencies adopted metric meter and kilogram measurements and standards. In 1866, the U.S. Congress first authorized the use of the metric system. Although internal progress is halting at best, the United States is one of the 17 original signers of the treaty establishing the International Bureau of Weights and Measures that was intended to provide worldwide metric standards. Most Americans do not know, for example, that since 1893, the units of distance (foot, yard), weight (pound), and volume (quart), have been officially defined in terms of their relation to the metric meter and kilogram.

After the modernization and international expansion of the metric system in the 1960s and 1970s following adoption of the SI, the United States soon stood alone among modern industrialized nations in failing to make full conversion. The English system was abandoned by the English as early as 1965 as part of Great Britain's integration into the European Common Market (a forerunner of the modern European Union) and countries such as Canada completed massive metrification efforts throughout the 1970s.

Following Congressional resolutions and studies that recommended U.S. conversion to the metric system by 1980, an effort toward voluntary conversion began with the 1975 Metric Conversion Act that established a subsequently short-lived U.S. Metric Board. The

American public simply refused to embrace and use metric standards.

It was not until 1988 the Congress once again tried to spur metric conversion with the Omnibus Trade and Competitiveness Act of 1988. The Act specified that metric measurements are to be considered the "preferred system of weights and measures for U.S. trade and commerce." The Act also specified that federal agencies use the metric measurements in the course of their business.

Regardless of the efforts of leaders in science and industry, early into the twenty-first century, U.S. progress remains spotty and slow. However, the demands of global commerce and the economic disadvantages of the use of non-metric measurements provide an increasingly powerful incentive for U.S. metrification.

Although the SI is the internationally accepted system, elements of the English system of measurement continue in use for specialized purposes throughout the world. All flight navigation, for example, is expressed in terms of feet, not meters. As a consequence, it is still necessary for a mathematically literate person to be able to perform conversion from one system of measurement to the other.

Real-life Applications

There are more than 50 officially recognized SI units for various scientific quantities. Given all possible combinations there are millions of possible conversions possible. All of these require various conversion factors. However, in addition to metric conversions, a wide range of conversions are used in everyday situations—from conversion of kitchen measurements in recipes to the ability to convert mathematical data into representative data found in charts, graphs, and various descriptive systems.

Historical Conversions

Historians and archaeologists are often called upon to interpret text and artifacts depicting ancient systems of measurement. To make a realistic assessment of evidence from the past they must be able to convert the ancient measurements into modern equivalents.

For example, the Renaissance Italian artist, Leonardo da Vinci used a unit of measure he termed a *braccio* (English: arm) in composing many of his works. In Florence (Italian: Firenze) *braccio* equaled two *palmi* (English: palms). However, historians have noted that the use of such terms and units was distinctly regional and that various conversion factors must be used to compare drawings and manuscripts. In Florence, a *braccio* equaled about 23 in. (58 cm), but in other regions (or among different professional classes) the *braccio* was several inches shorter. In Rome, the *piede* (English: foot) measured near it modern equivalent of 12 in. (30 cm) but measured up to 17 in. (34 cm) in Northern Italy.

Conversion of Temperature Units

Temperature can be expressed as units of Celsius, Fahrenheit, Kelvin, Rankin, and Réaumur.

The metric unit of temperature is the degree Celsius (°C), which replaces the English system's degree Fahrenheit (°F). In the scientists' SI, the fundamental unit of temperature is actually the kelvin (K). But the kelvin and the degree Celsius are exactly the same size: 1.8 times as large as the degree Fahrenheit. One cannot convert between Celsius and Fahrenheit simply by multiplying or dividing by 1.8, however, because the scales start at different places. That is, their zero-degree marks have been set at different temperatures.

The measurement of thermal energy involves indirect measurement of the molecular kinetic energies of a substance. Rather than providing an absolute measure of molecular kinetic energy, thermal measurements are designed to determine differences that result from work done on, or by, a substance (e.g., heat added to, or removed from, a substance). Temperature differences correspond to changes in thermal energy states, and there are several analytic methods used to measure differences in thermal energy via measurement of temperature. When dealing with the terminology associated with the measurement of thermal energy, one must be mindful that there is no actual substance termed "energy" and no actual substance termed "heat." Accordingly, when speaking of energy "transfer" or heat "flow" one is actually referring to changes in functions of state that can only be raised or lowered within a body or system. Neither energy or heat can really be "transferred" or "flow."

In thermodynamics, temperature is directly related to the average kinetic energy of a system due to the agitation of its constituent particles. In practical terms, temperature measures heat and heat measures the thermal energy of a system.

In meteorological systems, for example, temperature (as an indirect measure of heat energy) reflects the level of sensible thermal energy of the atmosphere. Such measurements use thermometers and are expressed on a given temperature scale, usually Fahrenheit or Celsius.

The common glass thermometer containing either mercury or alcohol uses the property of thermal expansion of the respective fluid as an indirect measure of the increase or decrease in the thermal energy of a body or system. Other types of thermometers utilize properties such as electrical resistance, magnetic susceptibility, or light emission to measure temperature.

Electrical thermometers (e.g., thermoprobes, thermistor, thermocouples, etc.) relate changes in electrical properties (e.g., resistivity) to changes in temperature are extensively used in scientific research and industrial engineering.

Because energy is commonly defined as the ability to do work, the thermal energy of a system is directly related to a system's ability to translate heat energy into work. Correspondingly, the measurement of the thermal energy of a system must be interpreted as the measurement of the changes in the ability of a system or body to do work. Absolute zero Kelvin—notice that Kelvin is not expressed as "degrees Kelvin"—(−459.69°F, −273.16°C, 0°R on the Rakine scale)—is the lowest temperature theoretically possible. At absolute zero there is a minimum of vibratory motion (not an absence of motion) and, by definition, no work can be done by a system on its surrounding environment. In this regard, such a system (although not motionless) would be said to have zero thermal energy.

In 1714, the German physicist Daniel Gabriel Fahrenheit (1686–1736) created a thermometer using liquid mercury. Mercury has a uniform volume change with temperature, a lower freezing point and higher boiling point than water, and does not wet glass. Mercury thermometers made possible the development of reproducible temperature scales and quantitative temperature measurement. Fahrenheit first chose the name "degree" (German: grad) for his unit of temperature. Then, to fix the size of a degree (°), he decided that it should be of such size that there are exactly 180° between the temperature at which water freezes and the temperature at which water boils. (180 is a "good" number because it is divisible by one and by 16 other whole numbers. That is why 360, or 2 × 180, which is even better, was originally chosen as the number of "degrees" into which to divide a circle.) Fahrenheit now had a size for his degree of temperature, but no standard reference values. Where should the freezing and boiling points of water fall on the scale? He eventually decided to fix zero at the coldest temperature that he could make in his laboratory by mixing ice with various salts that make it colder. (Salts, when mixed with cold ice, lower the melting point of ice, so that when it is melting it is at a lower temperature than usual.) When he set his zero at that point, the normal freezing point of water turned out to be 32° higher. Adding 180 to 32 gave 212°, which he used for the normal boiling point of water. Thus, freezing water falls at 32° and boiling water falls at 212° on the Fahrenheit scale. The normal temperature of a human being is about 99°.

In 1742, the noted Swedish astronomer Anders Celsius (1701–1744), professor of astronomy at the University of Uppsala (Sweden), proposed the temperature scale which now bears his name, although for many years it was called the centigrade scale. As with the Fahrenheit scale, the reference points were the normal freezing and normal boiling points of water, but he set them to be 100° apart instead of 180. Because the boiling point and, to a lesser extent, freezing point of a liquid depend on the atmospheric pressure, the pressure must be specified: "normal" means the freezing and boiling points when the atmospheric pressure is exactly one atmosphere. These points are convenient because they are easily attained and highly reproducible. Interestingly, Celsius at first set boiling as zero and freezing as 100, but this was reversed in 1750 by the physicist Martin Strömer, Celsius's successor at Uppsala.

Defined in this way, a Celsius degree (°C) is 1/100 of the temperature difference between the normal boiling and freezing points of water. Because the difference between these two points on the Fahrenheit scale is 180°F, a Celsius degree is 1.8 times (or 9/5) larger than a Fahrenheit degree. You cannot convert between Fahrenheit and Celsius temperatures simply by multiplying by 1.8, however, because their zeroes are at different places. That would be like trying to measure a table in both yards and meters, when the left-hand ends (the zero marks) of the yardstick and meter stick are not starting at the same place.

One method to convert temperature from Fahrenheit to Celsius or vice versa, is to first account for the differences in their zero points. This can be done very simply by (step 1) adding 40 to the temperature you want to convert. That is because -40° (40 below zero) happens to come out at the same temperature on both scales, so adding 40 gets them both up to a comparable point: zero. Then (step 2) you can multiply by 1.8 (9/5) convert Celsius to Fahrenheit or divide by 1.8 (9/5) to convert Fahrenheit to Celsius to account for the difference in degree size, and finally (step 3) subtract the 40° originally added.

WEATHER FORECASTING

An understanding of the daily weather forecast, especially in areas outside the United States requires the ability to convert temperatures between Celsius and Fahrenheit temperature scales. The standard conversion from Fahrenheit to Celsius is expressed as °C = (°F −32) / 1.8.

Accordingly a 72°F expected high temperature equates to approximately 22.2°C.

COOKING OR BAKING TEMPERATURES

To convert a temperature used for cooking (the expected oven temperature) for an French recipe for baking bread one might be called on to convert °C to °F and that conversion is obtained via °F = (°C × 1.8) + 32. So if an oven should be set at 275 °C in France to produce a crispy baguette (the traditional French long an thin loaf of bread) then an oven calibrated in °F should be set to approximately 525°F (275°C × 1.8) + 32 = 527°F.

Canceling Units

Notice that we are performing simple conversions, without the formality of labeling the units that must cancel to make the transformation. In the above example regarding oven temperature, the conversion factor 1.8 really represents 1.8°F / 1°C, read as 1.8 degrees Celsius to 1 degree Fahrenheit. This allows the units to cancel (275°C × 1.8 °F / 1 °C) + 32°F = 527°F.

In the prior example related to weather, the factor reciprocal of the factor 1.8 is used in the conversion formula °C = (°F − 32) / 1.8 equals 1°C per 1.8 °F or 1°C / 1.8°F and so the °F cancels as 22.2°C = (72 − 32) °F / 1.8 °C / °F.

ABSOLUTE SYSTEMS

About 1787 the French physicist Jacques Charles (1746–1823) noted that a sample of gas at constant pressure regularly contracted by about 1/273 of its volume at 0°C for each Celsius degree drop in temperature. This suggests an interesting question: If a gas were cooled to 273° below zero, would its volume drop to zero? Would it just disappear? The answer is no, because most gases will condense to liquids long before such a low temperature is reached, and liquids behave quite differently from gases.

In 1848 William Thomson (1824–1907), later Lord Kelvin, suggested that it was not the volume, but the molecular translational energy, that would become zero at about –273°C, and that this temperature was therefore the lowest possible temperature. Thomson suggested a new and more sensible temperature scale that would have the lowest possible temperature—absolute zero—set as zero on this scale. He set the temperature units as identical in size to the Celsius degrees. Temperature units on Kelvin's scale are now known as Kelvins (abbreviation, K); the term, degree, and its symbol, °, are not used. Lord Kelvin's scale is called either the Kelvin scale or the absolute temperature scale. The normal freezing and boiling points of water on the Kelvin scale, then, are 273K and 373K, respectively, or, more accurately, 273.16K and 373.16K. To convert a Celsius temperature to Kelvin, just add 273.16.

Conversion of measurements in recipes if often necessary.
ALEN MACWEENEY/CORBIS.

The Kelvin scale is not the only absolute temperature scale. The Rankine scale, named for the Scottish engineer William Rankine (1820–1872), also has the lowest possible temperature set at zero. The size of the Rankine degree, however, is the same as that of the Fahrenheit degree. The Rankin temperature scale is rarely used today.

Absolute temperature scales have the advantage that the temperature on such a scale is directly proportional to the actual average molecular translational energy, the property that is measured by temperature. For example, if one object has twice the Kelvin temperature of another object, the molecules, or atoms, of the first object actually have twice the average molecular translational energy of the second. This is not true for the Celsius or Fahrenheit scales, because their zeroes do not represent zero energy. For this reason, the Kelvin scale is the only one that is used in scientific calculations.

A traffic sign near the U.S. border in Quebec. OWEN FRANKEN/CORBIS.

ARBITRARY SYSTEMS

On the Réaumur scale, almost forgotten except in parts of France, freezing is at 0 degrees, and the boiling point is at 80 as opposed to 100° Celsius, or 212° Fahrenheit. The gradation of temperature scales is, however, arbitrary.

Conversion of Distance Units

Distance conversions are common to hundreds of everyday tasks, from driving to measuring. Conversion factors for distance are uncomplicated and easily obtained from calculators and conversion tables (e.g., 1 inch = 2.54 centimeters, 1 yard = 0.9144 meter, and 1 mile = 1.6093 km).

The meter was originally defined in terms of Earth's size; it was supposed to be one ten-millionth of the distance from the equator to the North Pole, going straight through Paris. However, because Earth is subject to geological movements, this distance cannot be depended upon to remain the same forever. The modern meter, therefore, is defined in terms of how far light will travel in a given amount of time when traveling at—naturally—the speed of light. The speed of light in a vacuum is considered to be a fundamental constant of nature that will never change, no matter how the continents drift. The standard meter turns out to be 39.3701 inches.

10K and 5K walks and races (measuring 10 and 5 kilometers, properly abbreviated km, or 10,000 and 5,000 meters) are popular events, often used for local charitable fund raising and well as sports competition. A 10K race is about 6.21 miles and a 5K race is, of course, half that distance (about 3.11 miles, with rounding). One kilometer = .6214 mile and so 10,000 km × .6214 miles/km = 6.21 km.

Other units of measurement related to distance encountered include: Admiralty miles, angstroms, astronomical units, chains, fathoms, furlongs (still used in horse racing), hands, leagues, light years, links, mils (often used to measure paper thickness), nautical miles (with different U.K. and U.S. standards), parsecs, rods, Roman miles (*milia passuum*), Thous, and Unciae (Roman inches).

Conversion of Mass Units

The kilogram is the metric unit of mass, not weight. Mass is the fundamental measure of the amount of matter in an object. For example, the mass of an object will not change if you take it to the Moon, but it will weigh less—have less weight—when it lands on the Moon because the Moon's smaller gravitational force is pulling it down less strongly.

Regardless, in everyday terms on Earth, we often speak loosely about mass and weight as if they were the same thing. So you can feel free to "weigh" yourself (not "mass" yourself) in kilograms. Unfortunately, no absolutely unchangeable standard of mass has yet been found to standardize the kilogram on Earth. The kilogram is therefore defined as the mass of a certain bar of platinum-iridium alloy that has been maintained since 1889 at the International Bureau of Weights and Measures in Sèvres, France. The kilogram turns out to be approximately 2.2046 pounds.

To convert from the pound to the kilogram, for example, it is necessary to multiply the given quantity (in pounds) by the factor 0.45359237. A conversion in the reverse direction, from kilograms to pounds, involves multiplying the given quantity (in kilograms) by the factor 2.2046226.

For large masses, the metric ton is often used instead of the kilogram. A metric ton (often spelled tonne in other countries) is 1,000 kilograms. Because a kilogram is about 2.2 pounds, a metric ton is about 2,200 pounds—ten percent heavier than an American ton of 2,000 pounds.

Some remnants of English weights and measures still exist in popular culture. It is not uncommon to have weights of athletes in football (American soccer) and rugby matches quoted by commentators in terms of "stones." A stone is the equivalent of 14 pounds, so a 15-stone goalkeeper or rugby forward would weigh a formidable 210 pounds.

Other units of mass encountered include carats (used for measuring precious stones such as diamonds), drams, grains, hundredweights, livre, ounces (Troy), pennyweights, pfund, quarters, scruples, slus, and Zentners.

Conversion of Volume Units

For volume, the most common metric unit is not the cubic meter, which is generally too big to be useful in commerce, but the liter, which is one thousandth of a cubic meter. For even smaller volumes, the milliliter, one thousandth of a liter, is commonly used.

Other units of volume include acre-feet, acre-inches, barrels (used in the petroleum industry and equivalent to 42 U.S. gallons), bushels (both United States and United Kingdom), centiliters, cups (both U.S. and metric), dessertspoons (U.S., U.K., and metric, and in the U.S. about double the teaspoon in volume) fluid drams, pecks, pints, quarts, tablespoons, and teaspoons.

Units such as tablespoons and teaspoons are among the most common of hundreds of units related to cooking where units can be descriptive (e.g., a "pinch" of salt). Most cookbooks carry conversions factors for units described in the book.

In the United States, gasoline is sold and priced by the English gallon, but in Europe gasoline is sold and priced by the liter. The unsuspecting tourist may not take immediate notice at the great difference in price because roadside signs advertising the two can sometime be very similar. Aside from differences in currency value explained below, a price of $2.10 per gallon is far less than 1.30 € (Euros) per liter. There are more than 3.78 liters per gallon and so the price of 1.30 €/liter must be multiplied by 3.78 to arrive at a gallon equivalent cost of approximately 4.91 Euros per gallon.

Currency Conversion

The price difference in the above fuel purchase example is exacerbated (increased not for the better) by the need to convert the value of the two currencies involved. As of mid-2005, 1 Euro equaled $1.25 (in other words, it took $1.25 to purchase 1 Euro). And so the actual price of the fuel in the above example was 1.30 Euro/liter × 1.25 $/Euro = 1.625 $/liter and thus a gallon equivalent price of $6.14 per gallon (1.625 $/liter × 3.78 liter/gallon).

Although currency values (and thus conversion factors) can change rapidly—over the years between 2001 and 2005 the Euro went from being worth only about 75 U.S. cents to more than $1.30—such price differences for fuel are normal, because fuel in Europe is much more expensive than in the United States.

Non-standard Units of Conversion

Another often-used, non-standard metric unit is the hectare for land area. A hectare is 10,000 square meters and is equivalent to 0.4047 acre.

Other measurements of area include Ares, Dunams, Perches, Tatami, and Tsubo.

Key Terms

English system: A collection of measuring units that has developed haphazardly over many centuries and is now used almost exclusively in the United States and for certain specialized types of measurements.

Derived units: Units of measurements that can be obtained by multiplying or dividing various combinations of the nine basic SI units.

Kelvin: The International System (SI) unit of temperature. It is the same size as the degree Celsius.

Mass: A measure of the amount of matter in a sample of any substance. Mass does not depend on the strength of a planet's gravitational force, as does weight.

Matter: Any substance. Matter has mass and occupies space.

Metric system: A system of measurement developed in France in the 1790s.

Natural units: Units of measurement that are based on some obvious natural standard, such as the mass of an electron.

SI system: An abbreviation for Le Système International d'Unités, a system of weights and measures adopted in 1960 by the General Conference on Weights and Measures.

Temperature: A measure of the average kinetic energy of all the elementary particles in a sample of matter.

Conversion of Units of Time, an Exception to the Rule

The metric unit of time, the second, no longer depends on the wobbly rotation of Earth (1/86,400th of a day), because Earth is slowing down; with days keep getting a little longer as time passes. Thus, the second is now defined in terms of the vibrations of the cesium-133 atom. One second is defined as the amount of time it takes for a cesium-133 atom to vibrate 9,192,631,770 times. This may sound like a strange definition, but it is a superbly accurate way of fixing the standard size of the second, because the vibrations of atoms depend only on the nature of the atoms themselves, and cesium atoms will presumably continue to behave exactly like cesium atoms forever. The exact number of cesium vibrations was chosen to come out as close as possible to what was previously the most accurate value of the second.

Minutes are permitted to remain in the metric system for convenience or for historical reasons, even though they do not conform strictly to the rules. The minute, hour, and day, for example, are so customary that they are still defined in the metric system as 60 seconds, 60 minutes, and 24 hours—not as multiples of ten.

Where to Learn More

Books

Alder, Ken. *The Measure of All Things: The Seven Year Odyssey and Hidden Error that Transformed the World.* New York: Free Press, 2002.

Hebra, Alexius J. *Measure for Measure: The Story of Imperial, Metric, and Other Units.* Baltimore: Johns Hopkins University Press, 2003.

Periodicals

"The International System of Units (SI)." *United States Department of Commerce, National Institute of Standards and Technology, Special Publication* 330 (1991).

Web sites

Bartlett, David. *A Concise Reference Guide to the Metric System.* <http://www.bms.abdn.ac.uk/undergraduate/guidetounits.html> (2002).

Coordinate Systems

Overview

Coordinate systems are grids used to label unique points using a set of two or more numbers with respect to a system of axes. An axis is a one-dimensional figure, such as a line, with points that correspond to numbers and form the basis for measuring a space. This allows an exact position to be identified, and the numbers that are used to identify the position are called coordinates. One example of the use of coordinates is labeling locations on a map. Street maps of a town, or maps in train and bus stations allow an overview of areas that may be too difficult to navigate if all features of the area were to be shown. Without a coordinate system, these maps would represent no sense of scale or distance.

The most common use of coordinate systems is in navigation. This allows people who cannot see each other to track their positions via the exchange of coordinates. In a complex transport system, this allows all the components to work together by exchanging coordinates that reference a common coordinate system. An example is an aviation network, where air traffic control must constantly monitor and communicate the positions of aircraft with radar and over radio links. Without a coordinate system, it would be impossible to monitor distances between aircraft, predict flight times, and communicate direction or change of direction to aircraft pilots over the radio.

Fundamental Mathematical Concepts and Terms

DIMENSIONS OF A COORDINATE SYSTEM

Coordinate systems preserve information about distances between locations. This allows a path in space to be analyzed or areas and volumes to be calculated. For example, if a position coordinate at one point in time is known and the speed and direction are constant, it is possible to calculate what the position coordinate will be at some future time.

The number of unique axes needed for a coordinate system to work is equal to the number of unique dimensions of the space, and is written as a set of numbers (x,y,z). In ordinary day-to-day life, there are three unique directions, side-to-side, up and down, and backwards and forwards. It was the German-born American physicist Albert Einstein (1879–1955) who suggested that there is a fourth dimension of time. This suggestion led to

Einstein's famous theory of relativity. However, these effects are normally not visible unless the velocities are very close to the speed of light or there is a strong gravitational field. Therefore, the dimension of time is not usually used in geometric coordinate systems.

Sometimes it is sensible to reduce the number of dimensions used when constructing a coordinate system. An example is seen on a street map, which only uses two axes, (*x,y*). This is because changes in height are not important, and locations can be fixed in two of the three dimensions in which humans can move. In this case, a coordinate system based on a two-dimensional flat surface (a map) is the best system to use.

CHANGING BETWEEN COORDINATE SYSTEMS

Coordinate systems denote the exact location of positions in space. If two or more sets of coordinates are given, it is possible to calculate the distances and directions between them. To see this, consider two points on a street map that uses a two-dimensional Cartesian coordinate system. A line can be drawn between the two points that extend from a reference point, say a building where a friend is staying, located at (*a,b*) on the map, to the point where you are standing (*x,y*). This line has a length, called a magnitude, and a direction, which in this case is the angle made between the line and the *x* axis. In Cartesian coordinates, the magnitude is given by Pythagoras' theorem:

$$\text{Magnitude} = \sqrt{(x-\text{a})^2 + (y-\text{b})^2}$$

The angle that this line makes with the *x* axis moving anticlockwise is given by:

$$\text{Angle} = tan^{-1}\left(\frac{y-\text{b}}{x-\text{a}}\right)$$

If you were to walk toward your friend along the line, the magnitude would change, but the angle would not. If you were to walk in a circle around your friend, the angle would change, but the magnitude would not.

You may have noticed that the magnitude (radius of the circle around your friend) and the angle taken together form a coordinate in the polar coordinate system, (*radius, angle*). These equations are an example of how it is possible to convert between coordinate systems. The Cartesian coordinates of your position can be redefined as a polar coordinates. The reverse is also possible.

VECTORS

This example also leads to the concept of vectors. Vectors are used to record quantities that have a magnitude and a direction, such as wind speed and direction or the flow of liquids. Vectors record these quantities in a manner that simplifies analysis of the data, and vectors are visually useful as well. For example, consider wind speed and direction measured at many different coordinates. A map can be made with an arrow at each coordinate, where each arrow has a length and direction proportional to the measured speed and direction of the wind at that coordinate. With enough points, it should be possible just by looking at this map to see patterns these arrows create and hence, patterns in the wind data.

CHOOSING THE BEST COORDINATE SYSTEM

Coordinate systems can often be simplified further if the surface being mapped has some sort of symmetry, such as the rotational symmetry of a radar beam sweeping out a circular region around a ship. In this case, the coordinate system with axes that reflect this circular symmetry will often be simpler to use. Coordinates can be converted from one system to another, and this allows changing to the simplest coordinate system that best suits each particular situation.

CARTESIAN COORDINATE PLANE

A common use of the Cartesian coordinate system can be seen on street maps. These will quite often have a square grid shape over them. Along the sides of the square grid, numbers or letters run along the horizontal, bottom edge of the map and the other along the vertical, left hand side of the map. In this example, assume that both sides are labeled with numbers. These two sides are called the axes and for Cartesian coordinate systems, they are always at 90 degrees to each other.

By reading the values from these two axes, the location of any point on the map can be recorded. The values are taken from the horizontal *x* axis, and the vertical *y* axis. The value of the *x* axis increases with motion to the right along the horizontal axis, and the value of the *y* axis increases with motion up along the vertical axis.

By selecting a point somewhere on the map, two lines are drawn from the point that crosses both the *x* axis and *y* axis at 90 degrees. The values along the two axes can then be read to give coordinates. The exact opposite technique will define a point on the map from a pair of coordinates. Two lines drawn at 90 degrees to the *x* axis and *y* axis will locate a point on the map where the two lines cross.

The coordinates for a point on the map are often written as (x,y). The order of expressing the coordinates is important; if they are mixed up the wrong point will be defined on the map.

Figure 1 shows an example of a two-dimensional Cartesian coordinate system. In three dimensions, a Cartesian system is defined by three axes that are each at 90-degree angles to each other. There is some freedom in the way three axes in space can be represented, and an error could invalidate the coordinate system. The usual rule to avoid this is to use the right-handed coordinate system. If you hold out your right hand and stick your thumb in the air, this is the direction along the z axis. Next, point your index finger straight out, so that it is in line with your palm; this is the direction along the x axis. Finally, point your middle finger inwards, at 90 degrees to your index finger; this is the y axis. The fingers now point along the directions of increasing values of these axes. A point is now located in a similar way to two-dimensional coordinates. From a set of coordinates, written as (x,y,z), a point is located where three planes, drawn at 90-degree angles to these axes, all cross.

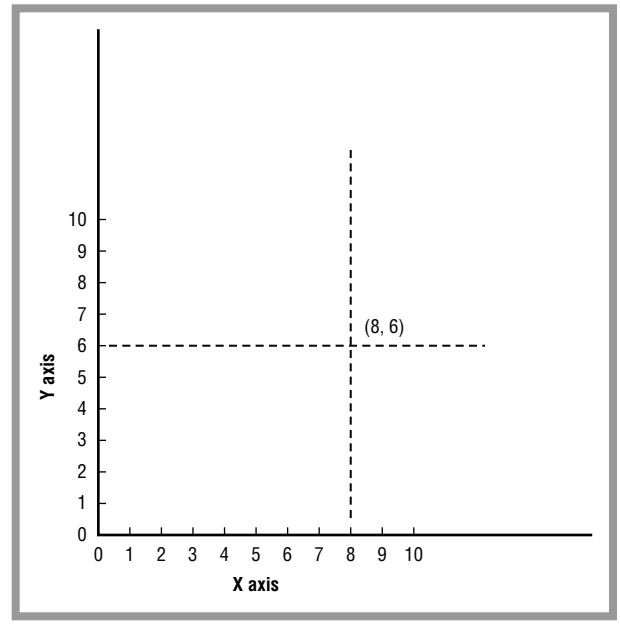


Figure 1: Rectangular coordinates.

POLAR COORDINATES

The polar coordinate system (see Figure 2) is another type of two-dimensional coordinate system that is based on rotational symmetry. The reason this system is useful is that many systems in nature exhibit rotational symmetry, and when expressed in these coordinates, they will often be simpler and more enlightening than using two-dimensional Cartesian coordinates.

The two coordinates used to define a point in this system are the radius and the polar angle. To understand this, imagine standing at the center of a round room that has the hours of a clock painted around the walls. Elsewhere in the room is a dot painted on the floor. The distance between you and the dot is the radius. The angle is a bit more involved. Standing facing 3 o'clock, the polar angle is given by the number of degrees you turn your head counterclockwise to face the dot. For example, if the dot is at the 12 o'clock mark, it has a polar angle of 90 degrees with respect to you; if it is at 9 o'clock, it has an angle of 180 degrees; and if it is at 6 o'clock, it has an angle of 270 degrees. The line at 0 degrees, the 3 o'clock mark, is defined to coincide with the horizontal, or the x axis in the Cartesian system.

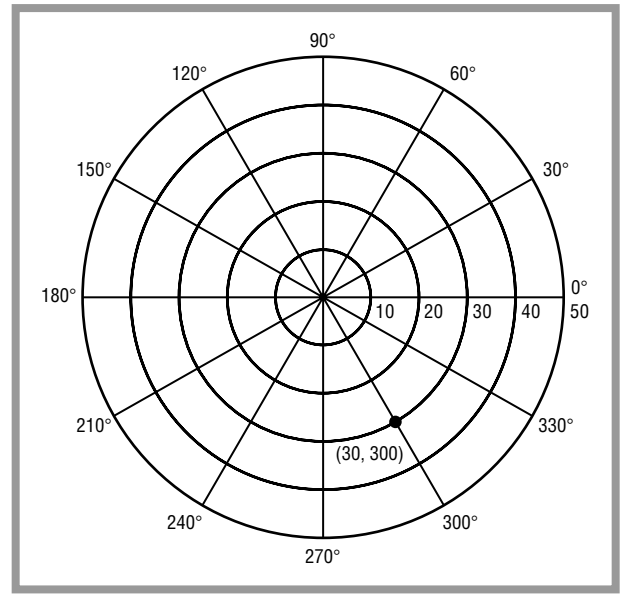


Figure 2: Polar coordinates.

A Brief History of Discovery and Development

Humans have been mapping their location and travels since the dawn of human history. Examples are seen throughout history, such as the mapping of land in the valley of the Nile in ancient Egypt, and recording journeys of global exploration such as those of the Spanish explorer Christopher Columbus (1451–1506) and others.

Today, the management of the world's natural and economic resources requires the availability of accurate and consistent geographic information. The methods for storing this data may have changed, with computer-based storage replacing paper maps, yet the underlying principals for ensuring compatibility has remained the same.

With coordinate systems, locations can be placed on maps and navigation can be achieved. Such systems allow a location to be unambiguously identified through a set of coordinates. In navigation, the usual coordinates in use are latitude and longitude, first introduced by the ancient Greek astronomer Hipparchus around 150 B.C.

Like so many mathematical ideas in history, coordinates may have existed in many forms before they were studied in their own right. French philosopher and mathematician René Descartes (1596–1650) introduced the use of coordinates for describing plane curves in a treatise published in 1637. Only the positive values of the x and y coordinates were considered, and the axes were not drawn. Instead, he was using what is now called the Cartesian coordinate system, named after him. The polar coordinate system was introduced later by the English mathematician and physicist Isaac Newton (1642–1727) around 1670. Nowadays, the use of coordinate systems is integral to the development and construction of modern technology and is the foundation for expressing modern mathematical ideas about the nature of the universe.

Real-life Applications

COORDINATE SYSTEMS USED FOR COMPUTER ANIMATION

Films makers and photographers use computers to manipulate images in a computer. Some common applications include photo manipulation, where images can be altered in an artistic manner, video morphing, where a computers morph an image into another image, and other special effects. Blue screen imaging is an effect where an actor acts standing in front of a screen, which is later replaced with an image. This would allow an actor dressed as Superman in front of a blue screen to later be seen flying over a town in the film, for example.

Leaps in computing power and storage have allowed animators to use computers to design and render breathtaking artistic works. Rendering is a process used to make computer animation look more lifelike. Some of these animations are works in their own right, and others can be combined with real life film to create lifelike computer generated effects.

All of these techniques require coordinate systems, as a computer's memory can only store an image as a sequence of numbers. Each set of coordinates will be associated with the position, velocity, color, texture, and other information of a particular point in the image. As an example, consider animating the figure of a dog in a cartoon. If the dog was featured in many scenes, it would be inefficient to redraw each movement of the dog. To simplify the animation, each part of the picture is split up into objects that can be animated individually. In this case, a coordinate system can be set up for each moving part of the dog.

For the finished animated picture, all the objects will be drawn together on some background image all at once, maybe with some objects rotated, shifted, or enlarged to refine the final effect. Vectors can be used to make this process more efficient and flexible. In two-dimensional animation and computer graphics design, this is often called vector graphics. In three-dimensional graphics, it is usually referred to as wire frame modeling.

COORDINATE SYSTEMS USED IN BOARD GAMES

Some games use boards that are divided up into squares. An example of this is chess, an ancient and sophisticated game that is played and studied widely. By defining a coordinate system on the board, the positions of the individual pieces can be located. Examples of this are found in books on the game and even in some newspapers, where rows of letters and numbers define the position and movements of the pieces. In this way, many famous games of chess have been recorded and a student of the game can replay them to learn tactics and strategies from masters of the game.

In computer chess simulators, the locations of the pieces have to be stored as coordinates as numbers in the computers memory. Once in the computer's memory, various algorithms calculate the movements of the pieces, which are then displayed on the computer screen.

Even without computers, if two chess players are separated by vast distances, the coordinate system allows the game to be played by the transmission of the coordinates of each move. There are many games of chess that have been played over amateur radio or by mail in this manner. In this case, the players can be separated by many thousands of miles and still play a game of chess.

PAPER MAPS OF THE WORLD

Assuming that the terrain one wishes to cross is flat, a coordinate system based on two dimensions and a Cartesian grid can be used for a paper map. This is suitable in shipping for maps of coastlines and maps of areas up to the size of large islands. However, the world is not flat, but curved, and for maps with areas larger than about 4 mi^2 (10 km^2), a Cartesian map of the surface will not be accurate.

One way to make an accurate map that covers most of the world on paper is to use a Mercator projection

(a two-dimensional map of the Earth's surface named for Gerhardus Mercator, the Flemish cartographer who first created it in 1569). This projection misses the North and South Poles, as well as the international date line. At the equator, the map is a good approximation of the Cartesian system, but because of Earth's curved shape, no two axes can perfectly represent its surface. Toward the poles, the image of the Earth's surface becomes more and more distorted. It is impossible to accurately project a spherical surface onto a flat sheet, as there is no way to cut the sphere up so that its sections can be rolled out flat. No matter what projection is used, flat paper maps of Earth's surface will always have some distortion due to the curved nature of Earth.

COMMERCIAL AVIATION

Coordinate systems allow a location to be transmitted over a radio link if two people have a map with a common coordinate system. Shipping is one example of this, but another important commercial use of coordinate systems is in aviation. In the skies, positions can be communicated as a series of coordinates verbally or electronically over radio links that allow many planes to be flown into or out of airports. In commercial aviation, there will often be many planes in the sky at one time coming in from all different directions toward an airport. At busy airports, sometimes there will not be enough runways to deal with all the traffic, and airplanes will often be put into a holding pattern while awaiting clearance to land. Positions of the aircraft are continually monitored by air traffic controllers with coordinates given both verbally by pilots and mechanically by radar.

As air traffic increases each year, it becomes more critical that coordinates and other information are relayed quickly and clearly. Air traffic controllers must make sure that coordinates are correct and understood clearly. Apart from all of the sophisticated technological safeguards, a simple misunderstanding of a spoken coordinate could be enough to cause a disaster. To avoid this, all commercial pilots must communicant in English, and flight terminology is common and standard across countries.

LONGITUDE AND JOHN HARRISON

In navigation, some point of reference is needed before a coordinate can be found. On a street map, a person could look for a street name or some other landmark to pinpoint their position. However, on the open seas and without fixed landmarks, it was not always simple for a ship to find a point of reference. To fix a position on Earth's surface requires two readings, called latitude and longitude. If the Earth is pictured as a circle, with the North Pole at the top and the South Pole at the bottom, and the ship is on the edge of the circle, the latitude is the angle between the ship, the center of the Earth, and the equator. Longitude can then be pictured as the circle when looking down from on top of the Earth, with the North Pole at the center of the circle. The angle between the ship and Greenwich, England is the longitude. Finding latitude is quite simple at sea using the angle between the horizon and the North Star or noon Sun. A device called a sextant was commonly used for this, but finding an accurate reading for longitude was more problematic.

Calculating longitude was a great problem in the naval age of the seventeenth and eighteenth century, and occupied some of the best scientific minds of the time. The British announced a prize of £20,000 for anyone who could solve the problem. It was finally solved by the invention of a non-pendulum clock that could kept accurate time at sea. It was invented by the visionary English clock maker John Harrison (1693–1776), who spent a great part of his life trying to construct a clock that was thought by many to be impossible with the technology of the time. It contained several technological developments that allowed it to work and keep time in the rough conditions at sea. During this time, John Harrison was constantly battling with the Royal Society, England's preeminent scientific organization. Ironically, while the members of the Royal Society were still debating if his clock really did work, it was already being used at sea for navigation by the navy. Eventually, after a long battle, John Harrison received the money and recognition he deserved. With the invention of this clock, calculating longitude at sea became simple. The clock is set to a standard time, taken as the time of Greenwich and called Greenwich Mean Time (GMT). If a person looks at the clock at noon, when the sun is directly overhead, and it reads 2 P.M., then two hours ago it was noon in Greenwich, as the sun rotates 360 degrees around the Earth every 24 hours. The equation is:

$$\frac{360^\circ}{24 \text{ hours}} \quad \begin{matrix}\text{2 hours}\\ \text{difference}\end{matrix} = \begin{matrix}\text{30° Longitude}\\ \text{from Greenwich}\end{matrix}$$

MODERN NAVIGATION AND GPS

In the twenty-first century, most navigation is based on the global positioning system (GPS). This is a network of 24 American satellites that orbit the Earth, allowing a position coordinate to be read off the screen of a special radio receiver anywhere on Earth, and is accurate to within 16.4 yd (15 m). Interestingly, this system requires use of a special coordinate system based on Einstein's theory of

Key Terms

Axis: Lines labeled with numbers that are used to locate a coordinate.

Coordinate: A set of two or more number or letters used to locate a point in space. For example, in 2 dimensions a coordinate is written as *(x,y)*.

Cartesian coordinate: A coordinate system were the axes are at 90 degrees to each other, with the *x* axis along the horizontal.

Dimension: The number of unique directions it is possible for a point to move in space. The world is normally thought of as having three. Flat surfaces have two dimensional and more advanced physical concepts require the use of more than three dimensions such as spa.

Polar angle: The angle between the line drawn from a point to the center of a circle and the *x* axis. The angle is taken by rotating counterclockwise from the *x* axis.

Polar coordinate: A two-dimensional coordinate system that is based on circular symmetry. It has two coordinates, the radius and the polar angle.

Radius: The distance from the center of a circle to its perimeter.

Vector: A quantity consisting of magnitude and direction, usually represented by an arrow whose length represents the magnitude and whose orientation in space represents the direction.

relativity called spacetime. In spacetime, time itself becomes a coordinate axis added to the normal three-dimensional world. The four-dimensional spacetime may seem strange, and the effects of it are far too small to be seen unless scientists or mathematicians are dealing with very high velocities or gravitational fields. However, the GPS satellites must give a very accurate time signal for the calculation of a coordinate. To do this, the satellites have small on-board atomic clocks. Relativistic effects from the high velocity of the satellites orbit relative to the Earth's surface distort this time signal and this distortion must be accounted for. If these effects were not taken into account, the resulting coordinates would be off by more than 6.2 miles (10 km) per day. This is all accomplished with an internal computer that returns the corrected map reading to the user.

3-D SYSTEMS ON ORDINANCE SURVEY MAPS

Some examples of three-dimensional coordinate systems can be found on ordinance survey maps. In this case, a two-dimensional Cartesian system is modified by the addition of lines to map height above sea level. These maps are used by surveyors and in sports, such as climbing and hiking, to map terrain with valleys and mountains. To define the height of the ground above sea level, two coordinates would not be enough. The basic map is a Cartesian system with a grid that gives two coordinates, but the third dimension for height is represented by curved lines drawn on the map. Each one of these lines represents a height in meters above sea level, giving the third dimension.

RADAR SYSTEMS AND POLAR COORDINATES

Modern radar systems are based on a device called a magnetron that produces a highly focused beam of microwaves. The beam can be rotated so that a radar operator can see all of a ship. A radar system that uses this method is seen on ships as a rotating parabolic aerial attached somewhere on top of the ship. This radar system is used to detect ships and other large solid objects in the sea, as the beam sweeps around the ship in a circular path. The radar screen will look like the familiar radar screen seen in movies, shaped as a round monitor with a line from the center sweeping around it in a circular path. Objects on the screen will show up as points as the beam sweeps over them.

The beam rotates in a two-dimensional fixed plane, so in order to locate objects, changes in height can be ignored, and a two-dimensional coordinate system can be used. The two-dimensional Cartesian coordinate system is not the best coordinate system to use in this case. Consider the operator's screen, for example. Although one might cover the round screen in a square mesh and put the round screen into a square box to draw the *x* and *y* axis, this would be impractical. The length from the center of the screen to a point to the edge of the round screen is constant, and is related to the maximum range the radar system can physically detect. As the edge of the

round screen is at maximum range, there would be areas dead areas between this and the square box used to define the Cartesian coordinate system. Another problem comes with the calculation of the distance and angles of objects in relation to the ship.

A better coordinate system to use in this example is the polar coordinate system, which reflects the circular nature of the sweeping beam. The radius axis is the distance along a line, drawn from the detected object to the center of the screen. The polar angle is measured between the horizontal line that crosses the center of the screen and the beam line. To draw a reference grid for the radius of this coordinate system, the screen is divided up into a number of concentric circles, or circles that get bigger with equal spacing, and are all centered at the screen center. Each of these circles is at a different fixed radius so the distance of the detected object can be read on the screen. A number of lines drawn at equal angles emanating from the center of the screen, like the spokes of a bicycle wheel, allow the polar angle to be read off, giving the angle between the ship and the detected object.

The center of the screen is always the location of the ship. If the radar operator sees a flash on the screen, the polar coordinate of the object is identified by the finding the circle and line that meet at the detected object. If each circle is labeled as 1km and each line labeled in 1-degree increments of angle, with the right hand side of the horizontal line representing the front of the ship, a polar coordinate made from the twentieth circle and the ninetieth line counter-clockwise from the horizontal instantly tells the radar operator that the object is 20 km away and 90 degrees to the right of the ship. More importantly, this information is read from the screen without using any mathematical conversion to find these figures, which would have been needed had a Cartesian system been used.

Where to Learn More

Books

Sobel, Dava, and William J. H. Andrewes. *The Illustrated Longitude*. New York: Walker & Company, 2003.

Web sites

Dana, Peter H. "Coordinate Systems Overview" *The Geographer's Craft*". University of Colorado. <http://www.colorado.edu/geography/gcraft/notes/coordsys/coordsys_f.html> (accessed March 18, 2005).

Stern, David P. "Navigation." <http://www-istp.gsfc.nasa.gov/stargaze/Snavigat.htm> (accessed March 18, 2005).

Decimals

Overview

Decimals can precisely indicate amounts, time speed to the hundredths or even thousandths of a second, precisely indicate the passage of time, accurately represent measurements of parameters that include weight, height, temperature and distance, and even help nab drivers who are speeding down the highway.

This article will consider decimals: what they are, how numbers are represented, and how decimals form a vital part of real-life math.

Fundamental Mathematical Concepts and Terms

The simplest way to answer this is visually: suppose that there are ten boxes on a table, as depicted in Figure 1.

Three of the boxes in Figure 1 are black in color and the remaining seven boxes are white. An ideal way to describe this relationship nonverbally is to use the language of math. A central part of a mathematical description can revolve around decimals. In order to write the preceding sentence using math language instead of words. The black colored boxes can be denoted as $1/10 + 1/10 + 1/10 = 3/10$. Another way to mathematically write the same information is in decimal form, expressed as 0.3.

This particular decimal consists of three components. The zero is in the ones column. Although other numbers are not present to the left of the zero, if they were, they would be in the familiar tens, hundreds, thousands, etc. columns. In other words, these numbers would be increasing from zero in $10\times$ increments. The number three is located immediately to the right of the period (the decimal point), in the column that depicts tenths (1/10).

If there was a number to the right of the three, that number would be in the hundredths (1/100) column. In the present example, 0.3, there are zero ones and three tenths. The number is pronounced as 'zero point three'.

Thus decimals can be seen as a short way of expressing certain types of fractions, namely those whose denominator are sums of powers of ten (tenths, hundredths, thousandths, etc.).

As an example, consider the number 8.53479. The number can be written in fractional form in terms of the place values of its various digits: $8.53479 = 8/1 + 5/10 + 3/100 + 4/1{,}000 + 7/10{,}000 + 9/1{,}000{,}000$. However, it is certainly a lot easier and more understandable to write this number in the decimal form (also called decimal

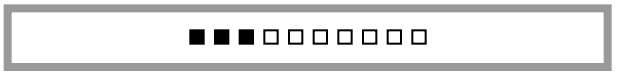


Figure 1.

notation) of 8.53479 than in the long and cumbersome fractional form.

A Brief History of Discovery and Development

Interestingly, although decimals are relatively new to numbering systems, base numbering systems like base 10 and base 60 have been around for thousands of years. In 1579, a book written by an Italian/French mathematician named François Viete contains a quote that argues for the use of the base 10 decimals (the tenths, hundredths and thousandths pattern seen above) instead of a more complex base 60 (sexagesimal) system that was then in vogue.

Viete argued, 'Sexagesimals and sixties are to be tested sparingly or never in mathematics, and thousandths and thousands, hundredths and hundreds, tenths and tens, and similar progressions, ascending and descending, are to be used frequently or exclusively.'

Just a few years later, in 1585, a book entitled *De Thiende* (The Tenth) popularized the concept and structure of decimals. However, the structure was a bit different than the decimals known today. The present day format of decimals came about in seventeenth century Scotland, courtesy of mathematician John Napier. It was Napier who introduced the decimal point as the boundary between the place values on ones and tenths. In some areas of the world a decimal comma is still used instead of a point.

Real-life Applications

As noted in the previous section, decimals numbers are easier to write and comprehend than numbers as represented in a fractional format, especially larger numbers. This ease of use and understanding has made decimals a centerpiece of disciplines including medicine, finance, and construction that call for the precise representation of distance, mass, and currency.

SCIENCE

In science, virtually all measurements are recorded and expressed as decimals. This accuracy is important to the scientific method, since it makes it possible for someone to repeat the reported experiments. Repetition of experiments and the resulting confirmation or refuting of the reported results is the cornerstone of science.

MEASUREMENT SYSTEMS

In countries that use the metric system, such as Canada and most of Europe, decimals predominate. Glancing at the digital thermostat might reveal a temperature of 68°F (20.17°C). A glance at the cereal box might reveal that a 1 cup (0.25 liter) serving of cereal contained 8.5 grams of protein and 2.7 grams of fat. A coffee bought at the local drive-through java emporium costs $3.00 plus a 15% tax (another $0.45).

Sports

There are many others examples of decimals in our everyday lives. Watch just about any sporting event in which timing of the game or the race is involved and a digital clock will inevitably be in use. Indeed, in track and field events like the 100-, 200- and 400-meter runs, the finish line clock is capable of measuring to the hundredths of a second. That is why a winning 100-meter time will be reported as 9.89 seconds, for example.

In the sport of baseball, a common practice for a team is to position one of their personnel in the stands to monitor the speed of the pitches thrown by the team's starting pitcher. Compiling this information can help the coach know at about what point in the game the pitcher starts to get tired and the velocity of his or her pitches begins to decrease. The timing device is used to record the speed of the pitches. This device is essentially the same as the one that police officers use to record the speed of vehicles zooming along a highway. These 'speed guns' display the speed digitally. So, when a coach sees the pitches drop to 75.5 miles per hour, or the police officer times a car moving at 80.3 miles per hour, action is likely to be taken.

GRADE POINT AVERAGE CALCULATIONS

Another example of one of the thousands of uses of decimals strikes motivating fear into the hearts of students, calculating their grade point average or GPA. The GPA is a cumulative score of the individual grades attained for the various courses taken. As high school seniors are well aware, universities, colleges and other institutions can place great emphasis on GPA when deciding on admittance of students.

Letter grade	Points
A	4.00
A−	3.67
B+	3.33
B	3.00
B−	2.67
C+	2.33
C	2.00
C−	1.67
D+	1.33
D	1.00
D−	0.67
F	0.0

Figure 2.

Bob		John	
A	4.00	A	4.00
B+	3.33	A	4.00
B	3.00	B+	3.33
C	2.00	B	3.00
C−	1.67	D	1.00

Figure 3.

GPA is based on the points that are assigned to a course. The points are usually based on a four-point grading scale similar to those in Figure 2.

In this example, Bob and John have received the following grades for the five courses taken: Bob received an A, B+, B, C, and a C−. John received two As, a B+, B, and a D. Using the grade point scale, the points for each of the courses is expressed in Figure 3.

In order to calculate the GPAs for the Bob and John, each student's individual scores are totaled and that number is divided by the number of courses. In other words, the average score is determined. Bob's GPA is (4.00 + 3.33 + 3.00 + 2.00 + 1.67) / 5, or 2.80. John's GPA is (4.00 + 4.00 + 3.33 + 3.00 + 1.00) / 5, or 3.066.

Where to Learn More

Books

De Francisco, C., and M. Burns. *Teaching Arithmetic: Lessons for Decimals and percents, Grades 5-6.* Sausalito: Math Solutions Publications, 2002.

Mitchell, C. *Funtastic Math! Decimals and Fractions.* New York: Scholastic, 1999.

Schwartz, D.M. *On Beyond a Million: An Amazing Math Journey.* New York: Dragonfly Books, 2001.

Web sites

BMCC Math Tutorials "Introduction to Decimals." <http://www.bmcc.org/nish/MathTutorials/Decimals/> (October 30, 2004).

Demographics

Overview

Demographics is the mathematical study of populations, and groups within populations.

Demographics uses characteristics of a population to develop policies to serve the people, to guide the development and marketing of products that will be popular, to conduct surveys that reveal opinions and how these opinions vary among various sectors of those surveyed, and of continuing news interest, to analyze polls and results related to elections.

Math lies at the heart of demographics, in the methods used to assemble information that is accurate and representative of the population. Without the accuracy and precision that mathematics brings to the enterprise, the demographic analysis will not provide meaningful information.

But demographics is not entirely concerned with math. Because demographics is also concerned with factors like cultural characteristics and social views, factors such as how people think about the issue at hand are also measured. Or, even less precisely, demographics can be concerned with how people 'feel' about something. These sorts of factors are more difficult to put into numbers and they are described as being qualitative (measuring quality) as opposed to quantitative (measuring an amount). Qualitative and quantitative aspects are often combined to form a 'demographic profile.'

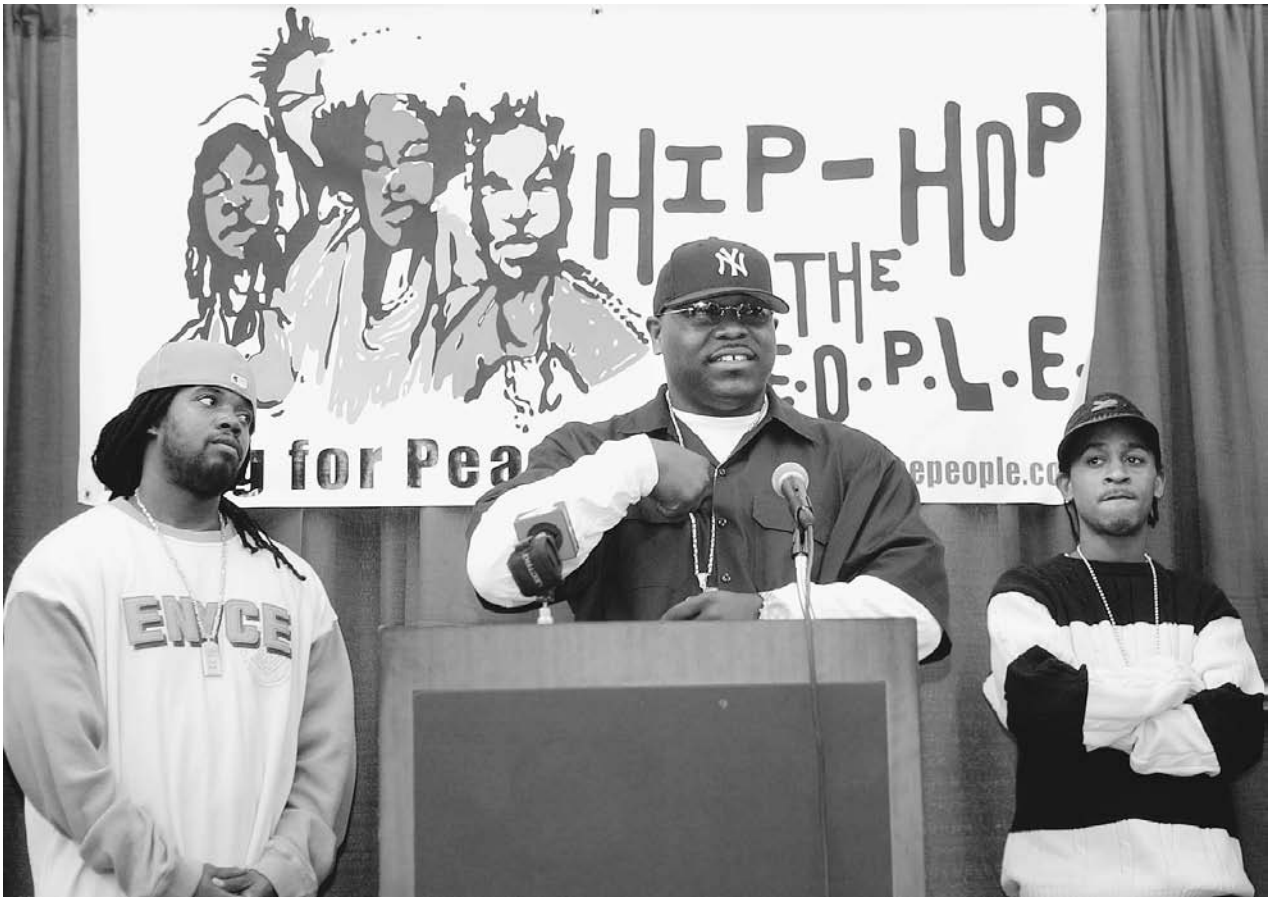
Some of the mathematical operations that can be useful in the analysis of demographic information include the mean (the average of a set of numbers that is determined by adding some aspect of those numbers and dividing by some aspect of the numbers), the median (the value that is in the middle of a range of values) and the distribution (the real or theoretical chances of occurrence of a set of values, usually patterned with the most frequently-occurring values in the middle with less frequently-occurring values tailing off in either direction.)

Demographic information can be very powerful. It can reveal previously unrecognized aspects of a population and can be used to predict future trends. Part of the reliability of the demographic information comes from the mathematical operations used to derive the data.

Real-life Applications

ELECTION ANALYSIS

The analysis of the 2004 general election (also called the Presidential election) in the United States offers an example of the use of demographics to analyze the voting patterns. By asking people questions about their beliefs and opinions on a variety of issues, and by utilizing databases



Artists (such as hip hop artists jace, Buckshot, and Freddie Foxxx, shown here) and other activists use demographics to identify specific areas and populations where advertising and money will be most effective. AP/WIDE WORLD PHOTOS. REPRODUCED BY PERMISSION.

that yield information on aspects such as age, gender, and income (more on this sort of information is presented below), a more complete picture can be built of the characteristics of those who voted for a certain candidate.

For example, exit polls (asking people questions after they have voted) were used to determine voter preferences and what issues were important in deciding how to cast votes in various races.

These characteristics can be considered along with information on employment, geographic residence, homeowner status, and other factors, to build up a profile of a 'typical' person who will vote for a particular politician.

These demographic patterns were known beforehand to campaign organizers, who conducted their own surveys of the public. So, aware of the characteristics of a certain segment of the population and the percentage of total voters who fit this demographic, candidates target specific groups with specific messages and promises.

CENSUS

Many countries periodically undergo a process known as a census. Essentially, a census is an organized gathering of information about the adult population of the country. Citizens and other eligible residents of the country complete a form or participate in an interview. Many questions are asked in a census. Example categories include age, gender, employment status, income range, educational background, marital status, number of dependents, ethnic background, place of residence (both geographically and in terms of whether a residence is owned or rented), history of residence change, and record of military service.

These categories of information can be analyzed to provide details of the characteristics of the population, and the proportions of the populations that make up each of the characteristic groups.

The demographic information in a census is used by governments to develop policies that will hopefully best

Cohort	Dates of birth	Events	Example characteristics
Great Depression	1912–1921	Depression, high unemployment, hard times	Need for financial security and comfort, Conservative
World War II	1922–1927	War, women working, a common enemy	The common good, patriotism, teamwork
Generation X	1965–1976	Space disasters, AIDS, safe sex, Berlin wall	Need for emotional security and independence, importance of money
Generation N	1977–present	September 11, Iraq wars, Internet	Need for physical safety, patriotism, increased fear, comfortable with change

Table 1.

serve their constituents. As well, the information represents a wonderful database for marketers to sell their wares. For example, it would not make sense for car company to target a region of high unemployment as a market for its top-of-the-line luxury car.

Demographics and the Marketplace

Demographics such as contained in a census have long been a tool of those who make and sell products. Knowing the characteristics, likes and dislikes of the buying public is obviously important when trying to sell a product.

The baby boom that occurred during the 1950s and 1960s provides a prime example of an identified demographic group. The increased birth rate in North America during those decades will have a number of effects that have and will continue to ripple through the ensuing decades. In the first few years, there was an increased demand for products to do with infants (baby food, diapers). Savvy entrepreneurs took advantage of the knowledge that an increasing number of new parents identified strongly with environmental protection to market organic baby foods and re-popularize nondisposable diapers. In the following few years as infants became youngsters, adolescents and young adults there was a succession of increased demands for children's toys and clothes, better educational facilities, housing and furniture. In the last decade, as the baby boomers have reached middle age, there has been an increased demand for certain types of vehicles such as SUVs, for health clubs and weight loss centers to help trim sagging waistlines, and for expertise in investment help as retirement draws closer. In the coming decades, as the baby boomers become infirmed, there will be a demand for more health-care services and funeral services.

Baby boomers came into the world at about the same time and, as they age, experience similar things and have similar demands. This generation is a perfect example of what was termed, way back in the 1920s, a 'generational cohort.' The designation has roots in mathematics. In statistical analysis, it can be advantageous and more meaningful to group items in cohorts that are similar in whatever aspect(s) is being studied. Historic examples of other demographic cohorts, and their associated characteristics, are given in Table 1.

GEOGRAPHIC INFORMATION SYSTEM TECHNOLOGY

Geographic information system (GIS) technology is the use of computers and computer databases to assemble information that have a geographical component. The information can come from reports, topographical maps that display elevation, land use maps, photographs, and satellite images of an area.

Knowledge of the geography can be combined with other data including information on age, gender, employment, health, and other aspects that are collected in a census, and data collected from other surveys. The aim is to provide a more complete picture of a region, in which demographic characteristics can be related to geographical features.

As an example, combining GIS data with population information could reveal that there is a higher incidence of fatal diseases in rural and mountainous areas. This could help health care providers in designing better ambulance service or telephone-based health advice.

The analysis and interpretation of geographic information can be a mathematical process. Equations can be applied to images to help sort out background detail from the more relevant information. Data can be statistically analyzed to reveal important associations between various data groups.

Where to Learn More

Books

Foote, D.K., and D. Stoffman. *Boom Bust & Echo: Profiting from the Demographic Shift in the 21st Century*. Toronto: Stoddart, 2000.

Rowland, D.T. *Demographic Methods and Concepts*. New York: Oxford University Press, 2003.

Wallace, P. *Agequake: Riding the Demographic Rollercoaster Shaking Business, Finance, and Our World*. London: Nicholas Brealey Publishing, 2001.

Discrete Mathematics

Overview

Discrete mathematics includes all types of math that deal with discrete objects, that is, things that are distinct, unconnected, or step-by-step in nature. For example, the natural numbers 0, 1, 2, 3, 4, . . . are discrete, and counting is a discrete process.

The increasing use of computers in science, engineering, mathematics, and daily life has led to fast growth in discrete mathematics over the last 30 years or so. Digital computers store numbers, words, and images in discrete form, that is, as ones and zeroes, so designing and programming computers involves discrete mathematics. Today, discrete mathematics is basic to most areas of computer science, including operating systems, algorithms, security, cryptography, networking, and database searching. Discrete math also used in engineering, chemistry, biology, operations research (the scientific management of large systems of people, machines, money, and material), and in many other fields.

The counterpart of discrete is continuous. Something is continuous if it changes smoothly from one place (or time) to another. A flight of stairs is discrete; a ramp is continuous. Calculus, which studies the behavior of curves and irregularly shaped areas, is an example of a branch of mathematics that is concerned mostly with continuous rather than discrete objects. However, discrete and continuous mathematics often overlap or influence each other.

Fundamental Mathematical Concepts and Terms

LOGIC, SETS, AND FUNCTIONS

The foundation of discrete mathematics is the study of logic, statements, sets, and functions.

Logic is the study of the rules of thinking. It helps mathematicians distinguish trains of thought that are valid (correct throughout) from ones that contain hidden errors. Logic can be considered a form of discrete mathematics because the steps in a logical train of thought occur one at a time, that is, in a discrete way.

Statements are claims about the way things are. True statements obey the rules of logic, and false statements break them. The statement $1 + 1 = 2$ is true, and the statement $1 + 1 = 5$ is false.

Set theory studies the ways in which statements, facts, or objects can be arranged into groups. These groups are called sets. For example, the two numbers 0 and 1 can be grouped into a set. Logic governs the making of statements about sets.

Finally, a function is a rule that connects the objects in one set with those in another. A function can be defined by a sentence like "Adding 1 to every number in set A gives a number in set B," by a formula like $f(x) = x + 1$, or by a computer program. Functions are basic to both discrete and non-discrete mathematics; they are the language that mathematics uses to describe changes and relationships.

BOOLEAN ALGEBRA

In 1854 the English mathematician George Boole (1815–1864) published a book called *The Laws of Thought*. In it, he described the rules for doing mathematics with the numbers 0 and 1. These are the rules of the type of discrete mathematics called Boolean algebra. In Boolean algebra, every operation (such as addition or multiplication) must give a result that is still a 0 or a 1. In ordinary arithmetic, $1 + 1 = 2$; but in Boolean algebra, $1 + 1 = 1$. Ordinary arithmetic can be approximately translated into Boolean algebra, which is how computers and calculators are programmed to do math.

NUMBER THEORY

Number theory is the study of integers, which are the counting numbers 0, 1, 2, 3, . . . and their negatives, –1, –2, –3, and so on. Numbers that cannot be written as whole numbers, like 1/2, 3/4, or 5.6, are not integers.

But how can there be a whole field of study devoted to something so simple? When you've counted 1, 2, 3, what else is there to say? Plenty, as it turns out. Prime numbers, for example, are one of the main interests of number theory. An integer is a prime number if it cannot be evenly divided by any number smaller than itself except 1. (Any number divided by 1 is just itself.) 4 is not a prime number because it can be evenly divided by 2, but 5 is prime because it cannot be evenly divided by 2, 3, or 4. One sure way to tell whether a number is prime is to try to divide it by every positive integer smaller than itself; if none of them divide the number evenly, it is a prime. Primes have surprisingly complicated and useful properties.

COMBINATORICS

Combinatorics is the mathematics of counting. Counting may also seem, at first glance, too simple to be a whole field of mathematics. When you count, you just point to each object to be counted and say "one, two, three . . ." until you run out of objects—right?

But this will not work if there is nothing to point at, or if there are too many things for one-by-one counting to be practical. This is often the case when we are trying to count not objects, but *arrangements* of objects, also called "permutations." For example, we might be designing a computer password system to serve as many as a billion users. We don't want to require extremely long passwords, because this might annoy users and drive them away. However, if we use passwords that are too short, there will not be enough passwords to go around. For example, if the passwords were only one letter long, there would only be 26 passwords (A, B, C, . . . Z). Would five letters be enough? We could answer such a question by writing down all possible five-letter combinations and then counting them, but this would take too long—remember, we want at least a billion passwords. Combinatorics answers questions like this efficiently. In the case of the five-letter password, one of combinatorics's simplest rules says that there are 26^5 or 11,881,400 possible passwords. This is not enough for a billion users. To allow for more than a billion passwords, we must use at least seven letters. Combinatorics enables us to count possibilities in more complicated situations, too, with many applications in computer science and other real-world fields.

PROBABILITY THEORY

Probability theory is the study of how likely things are to happen. For instance, when a fair coin is flipped, the probability of getting heads is 1/2 and the probability of getting tails is 1/2. Probability theory often uses continuous variables, but it is rooted in discrete mathematics because the "events" it deals with are separate or discrete.

Probability theory is used throughout science and business. Whenever we have to make a guess about the future—or about past events of which we cannot have certain knowledge—we must think in terms of probability. Corporations deciding how many items to manufacture and where to send them must decide what the market will probably want; gamblers and betters try to make the most probable bets; testing memory chips and other manufactured items for quality control is done using probability-based methods; and many computer tasks, such as searching a database for a particular name or other item, are treated by designers as "probabilistic" (random) processes. Combinatorics is used heavily in probability theory because to know how probable a particular event or group of events is, we need to know how many possible events there are. We calculate this using combinatorics.

ALGORITHMS

An algorithm is a set of instructions for solving a problem or performing a task. The problem may be mathematical—like deciding whether a particular

number is a prime number, or finding the square root of a number—or it may be non-mathematical, like baking a cake or finding a word in a computer database. Algorithms are used to tell computers how to do virtually everything that they do. When a digital camera focuses automatically, for example, its internal computer obeys an algorithm that tells it what part of the picture to focus on and how to know when that part is in focus. Large prime numbers, which are important for sending secret messages (cryptography), are found using algorithms. Algorithms are a part of discrete mathematics because the steps of an algorithm can be taken separately, one at a time.

CRYPTOGRAPHY

Cryptography is the science of making and reading secret (or "encrypted") messages—messages that look like completely random strings of symbols (letters, numbers, or bits). Cryptography is a sub-field of discrete mathematics because it deals with discrete (separate) symbols and words.

GRAPHS

In mathematics, graphs are drawings consisting of points (or circles) joined by lines. The circles are called nodes and the lines that join them are called edges. If you were to draw a five-pointed star, going directly from point to point with your pencil, and then put a circle at each point of the star, you would have drawn a graph with five nodes and five edges. Many real-world problems can be drawn as graphs. Nodes can stand for actual places (cities, say) connected by edges representing roads, railways, or telephone lines. Nodes can also be used to stand for states or conditions of a machine, with edges standing for possible changes from one state to another. For instance, two nodes connected by a single edge might stand for the ON and OFF states of a television, with the line between them standing for the fact that we can make the machine go from one state to another (turn it on and off).

MATRIX ALGEBRA

Matrix algebra gives the rules for handling matrices (plural of "matrix"). A matrix is a group of numbers or other symbols that have been arranged in a rectangular array, as if glued to the squares of a chessboard. Whenever we have a list of related mathematical equations (a "system of equations") and want to find a solution that satisfies all of them at once, we can write that list of equations as a matrix. Matrix algebra is used in computer programs designed to predict or mimic real-world events. Each number in computer memory can be treated as a number in a matrix, making it possible to solve large, difficult systems of equations efficiently.

Real-life Applications

SEARCHING THE WEB

You want to find something on the Web, so you call up the window of a favorite search engine such as AltaVista or Google and type in a word or phrase. In a few seconds or less, results appear—the first 10 or 20 out of what may be hundreds or even tens of thousands of matches, also called "hits." Somehow, in a fraction of a second, a computer (not yours, one belonging to the company that runs the search engine) has managed to comb through the contents of several billion Web pages to see which ones contain the word or words that you've entered. Each page may contain hundreds or thousands of words.

The search engine manages this trick by searching not the Web itself, but an index. A separate program is constantly "crawling" the Web, that is, automatically calling up hundreds of millions of Web pages. It then looks at all the words on each page and adds the words it finds to a large index or database along with information about how important each particular Web page might be. Such an index is huge—gigabytes or even terabytes (trillions of bytes)—but it is still far smaller than the Web itself. When you enter a word or phrase in the search engine, the engine searches the index. Structuring and searching large indexes and databases relies on the mathematics of graphs, especially that kind of graph called a "tree." Structuring an index as a tree makes searching it highly efficient. The result is that a search engine can dish up thousands of hits almost in almost as little time as it takes to get your request and send the results back.

COMPUTER DESIGN

When George Boole published his book *The Laws of Thought* in 1854, digital computers had not yet been thought of (though a few mechanical adding machines had been built). Boole's book, in which he laid out the rules of arithmetic using the simplest possible number system (0 and 1), was thought to be "pure" math, that is, math having no application to "reallife." But in 1938 the American mathematician Claude Shannon (1916–2001) showed that Boolean algebra could be used to design electrical circuits. It is easier and cheaper to build a circuit that represents 1 and 0 by switching itself on and off than to design a circuit that represents many numbers by switching between many in-between states. Today, all computer circuits are designed using Boolean algebra.

SHOPPING ONLINE AND PRIME NUMBERS

Cryptography is probably the most important application of number theory. It is, for example, basic to the functioning of the Internet. Without cryptography, millions of credit-card numbers could not be sent safely and automatically over the Internet every day. Whenever a web browser such as Explorer or Communicator announces that it has given you a "secure" connection after you have clicked on a link to make a credit-card purchase, a link using a type of encryption known as a "public-key cipher" is established. It is practically impossible for anyone (except, maybe, somebody at the National Security Agency, the United States government organization devoted to eavesdropping on communications and breaking codes) to read a message sent over such an Internet connection, even if they somehow manage to intercept the message somehow. Public-key ciphers depend on the fact that when very large two prime numbers are multiplied to give a third, larger number it is difficult—almost impossible, in practical terms—to discover the two primes from knowledge of their product (the large number made by multiplying them).

COMBINATORIAL CHEMISTRY

Many of the tastes and smells that we experience every day—whether in foods like french fries and gum, or wafting from toilet paper, shampoo, or makeup—are created in laboratories. Chemists are always looking for new taste and smell chemicals, "tastants and odorants" as they are called in the industry. One of the fastest-growing methods for finding new tastants and odorants, as well as drugs, pesticides, dyes, catalysts, and other chemicals, is combinatorial chemistry. In theory, scientists should be able to predict the properties of a complicated chemical just by looking at the shape and composition of its molecules. In practice, though, the only way to know how a complicated chemical will behave is to put it together and perform tests on it. Combinatorial chemistry, guided by the mathematics of combinatorics, creates small amounts of hundreds or thousands of similar chemicals all at once. These are then tested by computers at high speeds. Combinatorial chemistry has greatly speeded up the discovery of new drugs and other useful chemicals.

LOOKING INSIDE THE BODY WITH MATRICES

For about a hundred years, x-ray images of patients were taken by shining x rays (a type of high-energy light) through the body and capturing the shadows cast by body parts, especially bones, on a piece of photographic film. But in the 1980s, a new kind of x-ray came into being, called CAT scanning (for computerized axial tomography). In CAT scanning, a narrow x-ray beam, like the beam of a flashlight, is moved all around the patient in a circle. It is turned to point inward as it moves so that it shines through the patient crosswise from every possible angle. On the far side of the patient, an instrument records the power of the x rays shining through the patient. Where the x-ray beam meets more bone or other tissue that absorbs it, the beam is weaker on the far side of the patient. This process produces a long series of numbers (beam brightness measurements) that do not look anything like a picture of the inside of the patient's body—but using matrix algebra, a computer makes them into a clear, sharp "cross-section" image resembling what you would see if you could slice the patient in half. CAT scans show pictures of fine details inside the body that doctors could never see before. Other modern imaging methods, such as nuclear magnetic resonance imaging, also use matrix algebra.

FINDING NEW DRUGS WITH GRAPH THEORY

The chemical industry has for decades been building up databases that record the three-dimensional (3-D) structures of millions of molecules. The 3-D structure or shape of a molecule helps determine its medical properties. Researchers designing new drugs often know what shape they want a drug molecule to have in order to produce a certain effect in the body, but searching through millions of 3-D molecule records by calling up each one on a screen and looking at it is too slow. Instead, since the early 1990s drug designers have been using graph theory and algorithms to search for molecules with useful shapes. In this method, each molecule is represented as a graph, with atoms for nodes and chemical bonds for edges. Fast algorithms have been designed that look for matches between a "query graph" (the molecule the drug designer is looking for) and the graphs of the molecules in the database.

COUNTING JAGUARS USING PROBABILITY THEORY

Jaguars live in the jungles of Central and South America. In the wild, jaguars—like other predators—roam over vast areas, making it hard to know how many jaguars there are in a given area. Yet it is important to know how many jaguars there are in an area such as Kaa-Iya National Park in Bolivia, in order to know how best to protect them from extinction.

The probability theory of discrete events provides an answers. (Counting jaguars is discrete because jaguars are discrete.) Researchers set up "camera traps" in the forest

Key Terms

Boolean algebra: The algebra of logic. Named after English mathematician George Boole, who was the first to apply algebraic techniques to logical methodology. Boole showed that logical propositions and their connectives could be expressed in the language of set theory.

Combinatorics: The study of combining objects by various rules to create new arrangements of objects. The objects can be anything from points and numbers to apples and oranges. Combinatorics, like algebra, numerical analysis and topology, is a important branch of mathematics. Examples of combinatorial questions are whether we can make a certain arrangement, how many arrangements can be made, and what is the best arrangement for a set of objects. Combinatorics can be grouped into two categories. Enumeration, which is the study of counting and arranging objects, and graph theory, or the study of graphs. Combinatorics makes important contributions to fields such as computer science, operations research, probability theory, and cryptology.

Function: A mathematical relationship between two sets of real numbers. These sets of numbers are related to each other by a rule which assigns each value from one set to exactly one value in the other set. The standard notation for a function $y = f(x)$, developed in the 18th century, is read "y equals f of x." Other representations of functions include graphs and tables. Functions are classified by the types of rules which govern their relationships.

Logic: The study of the rules which underlie plausible reasoning in mathematics, science, law, and other discliplines.

Matrix: A rectangular array of variables or numbers, often shown with square brackets enclosing the array. Here "rectangular" means composed of columns of equal length, not two-dimensional. A matrix equation can represent a system of linear equations.

Prime number: Any number greater than 1 that can only be divided by 1 and itself.

Set: A collection of elements.

that automatically photograph jaguars as they pass by. Since the pattern of spots on each jaguar is unique, like a fingerprint, these photographs tell the researchers how many individual jaguars they have seen at each camera trap. The whole population of jaguars cannot be expected to walk past the cameras, however—it would cost too much to build that many camera traps—so a mathematical model is used instead, along with a method called "maximum likelihood estimation," that guesses what the most probable or likely total number of jaguars is based on the number that have been photographed. In 2004, biologists announced that using cameras and probability theory they estimated that there were about 1,000 jaguars in Kaa-Iya Park—more than they had thought, which is good news for this endangered species.

Where to Learn More

Books

Bogart, Kenneth P. *Discrete Mathematics.* Lexington, MA: D.C. Heath and Co., 1988.

Matousek, Jiri, and Jaroslav Nesetril. *Invitation to Discrete Mathematics.* Midsomer Norton, Avon, UK: Oxford University Press, 1999.

Rosen, Kenneth H. *Discrete Mathematics and Its Applications.* New York: WCB/McGraw-Hill, 1999.

Periodicals

Friedman, Richard R. "The suggestibility of children: scientific research and legal implications." *Cornell Law Review*, Nov. 1, 2000.

Web sites

National Institute of Standards and Technology. "Advanced Encryption Standard: Questions and Answers." Computer Resource Security Center. March 5, 2001. <http://csrc.nist.gov/encryption/aes/round2/aesfact.html> (June 16, 2004).

Thomas, Rachel. "Cat Count." Plus Magazine. May 31, 2004. <http://plus.maths.org/latestnews/may-aug04/jaguar/index.html> (Sep. 5, 2004).

Division

Fundamental Mathematical Concepts and Terms

Division is the inverse operation of multiplication, and is used to separate a set quantity into several smaller equal quantities. Simple division involves three quantities. The beginning value in a division problem is called the dividend, and the amount by which it is divided is labeled the divisor. The solution to a division equation is called a quotient, so a simple division equation takes the form of dividend / divisor = quotient.

Two symbols are used to signify a division operation. The commonly used division symbol (÷) is called an obelus, though this name is rarely used. The fraction line (/) is also used to signify division, and this symbol is called either a diagonal or a solidus. A fraction, written as one value separated from another by a solidus, is in reality a division equation that has not yet been evaluated. The upper (left-hand) value, or dividend, in a fraction is called the numerator, and the lower value or divisor is called the denominator. In cases where the dividend does not divide evenly by the divisor, the quantity left over after the division is termed a remainder.

A Brief History of Discovery and Development

As the inverse of multiplication, division probably developed around the same time it did. However, the need for complex division calculations emerged only fairly recently, and early applications of division were probably simple equations used to evenly divide and distribute quantities of tangible objects.

As the need for multi-digit calculations became more common, the process of long division was gradually refined, and relatively complex procedures were developed to deal with these increasingly challenging problems. The ancient Egyptians developed a repetitive, but effective, method for calculating long division solutions using only simple multiplication. In this process, the divisor is repeatedly doubled until the product is more than the original dividend; at the point this occurs, specific intermediate values from this process are added to find the solution. While this process works well for even division, additional complications arise when the initial division leaves a remainder, an outcome for which additional procedures were developed to approximate the resulting fractional result.

A second ancient method of division is attributed to the Hindus and goes by several names, including galley division, *batello* division, and scratch division. This

method, one of the most commonly used techniques prior to 1600, is well-suited to the abacus and counting table, but can quickly become confusing when done with pencil and paper due to the necessity of repeatedly crossing out or erasing numbers and replacing them with others. The basic methodology involves guessing a potential solution and replacing the existing value with this guess, then evaluating the outcome and making another guess. While tedious and needlessly complex in the twenty-first century, this method allowed lengthy equations to be evaluated many centuries before mathematics became a widely taught subject.

Ironically, one modern division technique is quite similar to the scratch method of repeatedly guessing and evaluating trial values. Trial division is a commonly used method of finding the prime factors of any number simply by evaluating different divisors or sets of divisors until a solution is found. While simple in application, this approach is computationally intensive, requiring hundreds or thousands of trials to produce a solution; however modern computers are ideally suited to such brute force methods, making this method useful in many situations.

Although powerful computers can solve virtually any division problem, equations with a divisor of 0 cannot be evaluated, and division by 0 produces an error message on pocket calculators, in spreadsheet formulas, and in computer programming applications. The reason for this error can be demonstrated by conducting a series of calculations which approach this value. For instance 1 / 1 = 1, and each reduction in the divisor produces a corresponding increase in the result. Accordingly, if a sequence of calculations is evaluated, we find that 1 / .001 = 1,000 and 1 / .0000001 = 10,000,000. As this progression continues and the divisor gets smaller and smaller (approaches 0), the result of the equation will grow larger and larger. If taken to its ultimate extreme, this process would produce the equation 1 / 0 = infinity (or, 1 / 0 = undefined). Because infinity is a symbolic value with little practical meaning, this progression explains a computer's inability to perform calculations in which infinity is the result. Given this impossibility, most computers respond with an error code, informing the user that dividing by 0 is illegal or impossible.

Real-life Applications

DIVISION AND DISTRIBUTION

Three boys have six cookies to eat; how do they determine the most equitable way to distribute the treats? For many people, a simple situation like this one will be their first exposure to one of the most common uses of division: distribution. In the case of boys and cookies, six cookies will divide evenly among three stomachs, and giving two cookies to each child will probably satisfy them all fairly well.

Of course if the boys are not all the same age and size, some may argue for a different distribution, with the largest boy arguing that he is entitled to three due to his larger appetite, or the smallest claiming he is due a larger share since his mother baked the cookies. While these childish disputes are relatively insignificant, they are different only in scale from the division disputes which take place daily in the world of commerce. For example, the 2004 major professional hockey season was cancelled, costing both players and owners millions of dollars in earnings, largely because the owners and the players could not agree how to equitably divide the profits of their joint venture. In this case, with the season cancelled, the income available to divide became 0, making the question of how to divide it much simpler.

Similar dilemmas arise when business ventures fail and companies declare bankruptcy. In such cases, a court will hear arguments from the company's management or their attorneys, and from claimants who have a legal right to receive some of the assets of the failed firm. In determining how to divide the assets, the judge is guided in part by a specific set of laws which govern bankruptcy and his or her own understanding of how strong each claimant's position is. In such cases, it is not unusual for stockholders to receive nothing, while bondholders and others receive twenty-five to seventy-five cents for each dollar they are owed.

Large corporations do not have an individual owner, per se, but are owned by their stockholders, a diverse group which can include individuals, pension funds and retirement plans, and other corporations. These owners are compensated for their investment by the payment of dividends, normally distributed on a quarterly schedule. A typical method of accounting for corporate earnings allows the firm to subtract all its costs of doing business, as well as the taxes it pays, to produce a final net earnings value. From these net earnings, the firm will take some funds to invest in future growth and other strategic goals; the remainder will typically be distributed to shareholders as a dividend, normally by paying an equal share to the owner of each share of common stock.

In the case of a large corporation such as Apple Computer, General Motors, or Wal-Mart, earnings for a given quarter could be hundreds of millions of dollars. However, the number of shares of stock is so large that a

firm's dividend, or the amount that it passes along to shareholders, is usually fairly small, often in the range of $1.00 to $2.00 per share. In order to determine the value of this payment to an investor, analysts frequently calculate a value known as dividend yield, which equates to the percent of the share price which the annual dividend equals. Dividend yield is simple to calculate: annual dividend is divided by stock share price to give yield. This value provides an estimate of the first year return an investor might make if he or she purchased shares of the firm. Dividends are typically not paid by firms which are losing money, or by rapidly growing start-up companies which prefer to reinvest earnings in their current business. Microsoft, a leading software company, grew rapidly through the 1980s and 1990s, however it did not pay its first regular dividend until 2004.

Budgeting, whether dealing with money or any other limited commodity, provides many opportunities to apply division. An annual income of $48,000 can be budgeted across 12 months using division, giving a monthly spendable income of $4,000; in practical terms, many expenses do not appear evenly across the months of the year, so a prudent budget will include some unspent money for unforeseen or irregularly timed expenses.

At the corporate level, high-flying companies often start with a large pot of investor cash and race to become profitable before the pot runs dry; these firms, or the analysts who watch them, frequently assess their progress using the graphically labeled "burn rate." A company's burn rate is simply the amount of cash which it is burning, or spending in the course of a month, and is found by subtracting the current month's cash reserves from last month's figure to find out how much has been spent. Once the burn rate is calculated, the company's total cash nest egg can be divided by this figure top project an expected lifetime for the company before it runs out of cash and goes out of business. In the case of a firm with a $45 million cash start and a $9 million burn rate, the equation is 45 / 9 = 5, giving the firm's managers five months to either turn a profit, find additional investors, or look for other employment. Many of the internet startup firms of the early twenty-first century had burn rates of several million dollars per month, and in most cases, they never reached profitability.

Rationing is another form of budgeting, in which scarce resources such as water or food are distributed at a slower rate in order to prolong the supply. Survivors of a shipwreck might determine that they have eighteen gallons of water to divide among three survivors, and simple division will tell them that six gallons per person is all they have. Similarly, these same survivors might also wonder how long their limited water supply will last; dividing each individual's supply of six gallons by each person's minimum consumption of one gallon per day provides the bad news that the water will last less than one week.

DIVISION AND COMPARISON

Division is frequently used to compare two items to each other using a relationship known as a ratio. For example, a researcher might wish to know how accessible medical care is in different parts of the country. Common sense would tell her that a large city like Chicago should have more doctors than a small town like Whiteland, Indiana, but how can she compare these two, since they are so different? One approach to making this comparison involves the use of ratios. Dividing the number of doctors by the number of residents gives a ratio of doctors to residents, such as one doctor for each 1,400 residents. For example, if Whiteland has 2,700 residents and 3 doctors, dividing these values would produce a ratio of 1 doctor for every 900 residents. Other locations would have different ratios; some small towns might have ratios of 1 doctor for every 5,000 or more residents, while an urban area with a teaching hospital or an affluent suburb might have one doctor for every 100 residents. Using this ratio calculation, the researcher will be able to compare any locale with any other. In most cases, the conclusion will be that residents in areas with a higher ratio of doctors will have better access to medical care.

In numerous settings, division can provide a measure of the relationship between two quantities. Environmental scientists wishing to examine the rate of deforestation could choose to assess the density of trees in an area by dividing the number of trees in a forest by the number of acres covered, producing a ratio of trees per acre, which can be tracked over time. Similar methods can be used to assess the number of people living in an area by counting the population in a city, then dividing by the number of square miles in the urban area to produce a measure of residents per square mile.

Speed is commonly expressed in terms of miles or kilometers per hour, a simple ratio of distance covered in a set quantity of time which can be calculated for a given trip by dividing the number of miles covered by the number of hours required for the trip. Chemical concentrations are frequently expressed in ratio form using the measure parts per million (PPM), or the number of particles of a given substance which would be found in a million particles of the combined substance. Home swimming pools typically use low levels of chlorine to keep algae and bacteria from growing in the water, and

the chlorine level in a swimming pool is typically measured using this ratio. While chlorine levels will vary depending on the weather and the number of swimmers, appropriate levels of 1 to 3 PPM will produce comfortable swimming and low maintenance requirements. At higher levels, chlorine stings the eyes of swimmers and tints hair, while lower levels are inadequate to control algae, causing the pool water to gradually shift from clear to green.

AVERAGES

Division is commonly used to produce averages. An average allows a large amount of data to be clearly and succinctly summarized in a single value. A simple average is found by adding the quantities involved and dividing by the number of items. If the daily temperatures for a week are 70, 72, 71, 75, 77, 71, and 45, the average is found by adding all the values, then dividing by seven, which is the number of data points included. In this case, the average would be 481 / 7 = 68.71, meaning the average temperature throughout the week was 68.71 degrees.

Because averages sum up a large amount of data, they frequently pay for this efficiency by obscuring individual values or trends in the data. In the temperature example above, a seasoned weather watcher would probably conclude from the raw data that a cold front blew into the region on the last day of the week, causing the temperature to drop rapidly, and significantly impacting the average. Recalculating the average for the first six days of the week produces an average temperature of 72.67, a value much closer to the temperatures actually experienced on the first six days of the week.

In this example, the accuracy of the average was improved by eliminating the final temperature reading from the calculation. Values which lie far from the rest, referred to as outliers, can reduce the accuracy of a simple average by skewing the results away from what it would otherwise be. For example, when evaluating average income in a city, one might find that 450 residents each earned an average of $40,000 per year while a single wealthy resident earned $9 million. While an estimate of $40,000 per resident would be quite accurate for a typical wage earner in this town, calculating the average income for the area gives a value of $59,866. This value provides a misleading picture of the area's income by suggesting that the average wage-earner in the area actually makes approximately 50% more than he actually does. In addition, no person in the city earns a salary anywhere close to this value, making its usefulness as a summary value suspect.

A second weakness of simple averages is that they sometimes produce values which are mathematically correct, but which actually provide a misleading picture of the underlying values. For example, a store owner might wish to assess the age of his customers in order to better serve their needs. To accomplish this, he assigns an employee to ask customers their ages and then calculate an average age. As soon as the manager looks at the average age of 45, he contemplates firing the employee, since he knows from observation that most of the customers in his store walk over from a nearby high school and a senior citizens center, and he has seen nobody near the age of 45. But further investigation reveals that the average is actually correct, since the ages of the customers were 14, 15, 16, 74, 75, and 76. In this case, where the total population is made up of two distinct groups, rather than a continuous distribution, the averaging process produces a solution which is mathematically correct, but practically misleading. In such cases, other measures of central tendency may be used, either alone or in concert with the average. One such measure, the mode, locates the most common or frequently occurring value in a distribution; this measure is useful when values are presented in whole numbers, such as ages. A second measure, the median value, locates the center-most value of a sequence. This assessment is useful for dividing a distribution in half, since it provides a line which lies below half the values in the distribution and above the other half.

Averages are often used to compare athletic performers and coaches. A baseball player's hitting is calculated by dividing his number of hits by his number of times at bat, though walks, fielder's choices, and other outcomes are counted differently than actual safe hits. This value, the player's batting average, tells how often a given hitter can expect to bat successfully in a given number of attempts. In the case of a batter who has hit fifteen times out of 100 attempts, the batting average would be 15 / 100 = .150. Batting averages are generally recorded to three decimal places and stated without the decimal point, meaning that these averages would typically be stated as "one-fifty."

Coaches are often assessed on their ratio of wins to attempts, or the number of wins, compared to the total number of games played. A successful coach might have compiled a record of 400 wins and 125 losses. This coach would have won 400 of his total 525 games, and dividing these values (multiplied by 100 to create a percentage) would produce a career winning percentage of 76%. Top-ranked coaches in professional basketball and football are frequently able to win far more games than they lose, producing high winning percentages.

In many cases, a simple average provides a useful summary of the values involved. But in some situations, one or more of the items being averaged is considered

more important than the others; in these cases, a weighted average may be used to more accurately assess the data. A weighted average assigns a factor, or weight, to each value in the data set, giving some values a larger influence on the final result.

Consider the case of a new car buyer who wishes to compare three vehicles. The buyer has decided that his major criteria are price, efficiency, size, and sportiness, and he proceeds to rate each of the three vehicles on each of these four criteria using a scale from 1 to 10. However, the buyer is on a limited budget, so financial considerations are more important to him than other factors. For this reason, he assigns a weight of 1.0 to the factors of size and sportiness, but a weight of 2.0 to price and a weight of 1.5 to efficiency, which impacts the long-term cost of ownership.

Once these weights are multiplied by the raw scores to produce weighted scores, the resulting average will give the factor of price twice the influence of size and sportiness, while efficiency will be one and one-half times as influential. Weighted averages can be extremely helpful as a decision-making tool by allowing specific factors to influence the final score more highly than others.

PRACTICAL USES OF DIVISION FOR STUDENTS

Students are frequently encouraged to learn good time management skills, that in many cases requires making projections to determine how long a project might take. For example, a student who must read a 30-page chapter and has allotted one hour to read it can divide sixty minutes by thirty pages to find that he can devote two minutes to each page if he wants to finish in the allotted time. He can also determine that if he chooses to spend his first thirty minutes watching television, he will then have only thirty minutes left for his assignment, giving him one minute per page.

Similar choices exist when meal time arrives. The student eating on a budget of $15.00 per day can easily figure out using division that her choices will determine both how well and how often she eats. For instance, if she keeps her meal costs to $5.00 a piece, division tells her that she can plan to eat three meals per day. However if her meal cost climbs, the number of meals will fall; at $6.00 per meal, the number of meals will be 15 / 6, = 2.5, meaning she can afford 2 full meals and a snack, or perhaps a soft drink at mid-afternoon.

When a semester ends, one particular average takes center stage: the grade point average, or GPA. GPA is an average that summarizes a student's grades for a single semester or an entire academic career. GPA is found by assigning a scale to the letter grades, most commonly with a grade of A earning 4 points, B earning 3, and so on. These grades are then averaged by adding the values and dividing by the number of courses taken (or more frequently by the number of credit hours earned), and this ratio is the GPA. A perfect GPA of 4.0 would mean a student had earned only grades of A during the entire academic career, while a GPA of 1.0 would mean the student's average grade was a D.

Calculations of GPA are important to colleges evaluating applicants and to insurance companies. Students with high GPAs have also been found to experience fewer automobile accidents, a relationship that some insurers exploit by offering good student discounts to these young drivers.

OTHER USES OF DIVISION

Measuring a nation's affluence can be difficult, particularly when trying to compare it to another country with different characteristics. Consumer affluence is often measured by examining the availability of consumer goods, such as automobiles, televisions, and housing. While these goods are available in virtually every economy on Earth, the relationship between an item's price and a consumer's hourly wages may make it unaffordable.

Consider an automobile which sells for $20,000 in the United States, and a comparable car which sells for $8,000 in a less-developed nation. Before concluding that the car is more affordable in the other country, it is necessary to compare these prices to the wages of the citizens who might purchase them. Assuming that the average hourly wage in the U.S. is $10.00, we can divide the purchase price of the car by this value to conclude that a typical U.S. worker will spend 2,000 hours at work in order to pay for the car (and in reality much more because that income does not include deductions for taxes, etc.).

In the case of the second country, low prices are accompanied by much lower wages, in this case an average hourly rate of $1.00. Dividing the $8,000 price of the car by the hourly wage tells us that a worker in this country will have to work 8,000 hours to buy the car, or four times as long as the American worker. If similar relationships hold for other consumer goods, one can reasonably conclude that the American worker will be able to purchase more consumer goods for a given number of hours worked, and will enjoy a higher standard of living.

Retailing is a competitive industry, and in addition to the challenge of facing their competitors, retailers often face problems caused when their own employees either steal or give away merchandise, a crime known as pilferage. In many cases, employees assume that because the

firm has such a large supply of hamburgers, pencils, or other items, a few will not be missed and will cause little harm. Unfortunately, this assumption is often incorrect.

Consider a young man working at a hamburger restaurant. After several months of work, he decides to do something nice for his best friend, so when the friend orders a meal, the young man simply does not charge him for his hamburger, which would normally sell for $1.65. What is the impact on the company of this theft? One way to determine the damage is to calculate how many hamburgers must now be sold in order to offset the loss caused by the employee's dishonesty. The company must first pay its owns costs before it earns any profits, and its expenses for rent, beef, wages, lettuce, and all its other costs come to $1.50 per hamburger, meaning that its net profit per sandwich is only 15 cents.

Because this 15 cents in profit is all that can be used to pay for the materials in the stolen burger, the company must divide the lost cost of the stolen burger by the profit earned on each additional burger in order to determine how many must be sold to recoup the loss. In this case, the $1.50 loss divided by 15 cents profit per sandwich means the restaurant must sell ten more burgers to make up the loss incurred on the stolen sandwich. These ratios are typical for most retailers, making employee theft a major threat to company profitability.

As digital cameras rapidly replace film cameras, users must determine what size memory card to purchase. One simple way to answer this question uses division to calculate how many photos will fit on each card. Imagine a photographer who wants to know how many shots he can take before he fills up his memory card and is forced to either download or delete shots. The memory card in this case is a 512 megabyte model, and the user's camera produces shots requiring 4 megabytes of storage apiece. Dividing the card capacity by the size per shot, the user finds that he will be able to store 128 digital pictures at a time.

Percentages are one of the most common methods of using division to represent and compare different values. For example, imagine that two extreme sports fans are arguing about which of their two favorite competitors is a better street luge racer. One racer has competed in 200 races and has won 150, while the other, who is in his rookie season, has only competed 38 times, of which he has won 30. Which racer is having a more successful career?

While many other aspects of the question could be argued, one simple comparison would be to calculate which racer has won more of his competitions. Obviously the first racer has won more races than the second has even entered, so how can these two performances be compared? Calculating a percentage allows both scores to be standardized for easy comparison.

A percentage expresses a ratio or division equation as if its divisor were 100. For example, in examining the human population of earth, we could easily determine that every living human being on the planet is breathing, meaning that the ratio of breathing humans to humans is around 7,000,000,000 / 7,000,000,000, or 1.0. Expressing this value as a percentage simply requires moving the decimal point two places to the right; thus 1.0 equals 100%, which is the percentage of living humans who are breathing.

In a similar manner, we could count and determine that of the 7 billion humans on Earth, about 3.4 billion are male, meaning that the ratio of males to humans is 3.4 / 7, and if we solved this equation we would find a decimal value of .486, or 48.6%. A similar process would tell us that the U.S. population of 300 million accounts for only 6% of the total world population.

Applying this percentage technique to our original question, we can calculate each racer's winning percentage in order to compare them. For the more experienced racer, 150 wins divided by 200 races produces a win ratio of 75%, meaning that in 100 races he can expect to win about 75. How does this compare to the newcomer's performance? Dividing his 30 wins by his 38 races gives a win ratio of 78.9%, meaning that in the same 100 races he would probably triumph in 79 of them, or about 4 more races than his competitor. While this comparison may not settle the debate over which extreme racer is better, it does provide a simple technique for comparing one racer's performance to another.

Many drivers get in a hurry, sometimes choosing to disobey posted speed limits in order to arrive at their destinations more quickly. What is the cost of this choice, assuming the driver is pulled over by a highway patrol officer? One way to assess the cost of this decision involves determining how much the driver paid for each mile over the speed limit he drove, and this can be found by dividing the fine by the number of miles over the limit. In this case, a fine of $75.00 for going 15 miles per hour over the limit would mean the driver paid $5.00 for each mile per hour that his speed exceeded the limit.

Car maintenance can extend an automobile's life; in particular, changing a car's oil regularly will improve its chance of a long life. Today's car owner faces many choices in the oil market, and division can help her determine the relative cost of each choice. Assuming this driver changes her own oil, she might wonder whether she should use standard motor oil, which costs $1.00 per quart, or synthetic oil, which costs $5.00 per quart.

Key Terms

Dividend: A mathematical term for the beginning value in a division equation, literally the quantity to be divided. Also a financial term referring to company earnings which are to be distributed to, or divided among, the firm's owners.

Percentage: From Latin for per centum meaning per hundred, a special type of ratio in which the second value is 100; used to represent the amount present with respect to the whole. Expressed as a percentage, the ratio times 100 (e.g., 78/100 = .78 and so .78 × 100 = 78%).

Stockholder: The partial owner of a public corporation, whose ownership is contained in one or more shares of stock. Also called a shareholder.

At first glance, synthetic oil appears far more expensive, since a typical oil change requires five quarts. However, the newest synthetic oils claim they will last a full year, while standard oil must be changed 3–4 times per year. How can a driver compare these options? A simple division equation allows a direct comparison of these two possibilities.

An oil change using five quarts of synthetic oil will cost $25.00, while each of the year's four changes using standard oil will cost $5.00 to drive the car for three months. Dividing the $25.00 cost of the synthetic change by four, we find an equivalent value for the synthetic change of $6.25, compared to $5.00 for standard oil. While this calculation demonstrates that synthetic oil does cost an average of about $1.25 more per oil change period (or about $5.00 more per year), this calculation does not take into account the other costs associated with the two options, such as the fact that one produces four times the quantity of used oil which must be disposed of, or that the time required to change the car's oil four times rather than one may be valuable to some owners.

Potential Applications

A traditional telephone system sends a person's voice as a continuous electrical signal over a pair of wires. But data sent over most digital data networks is actually divided into numerous small pieces, or packets, which are sent across the network independently, then reassembled at the destination. This method of dividing a message into numerous small pieces offers numerous advantages, including greater efficiency and much lower costs than the use of a dedicated line.

One of the most intriguing uses of this digital division process is the rapid emergence of VOIP (voice over internet protocol) telephone systems. The systems take the sound of a speaker's voice, break it into small packets of data, and send these packets across the Internet with all the other data packets traveling there; at their destination, the packets are reassembled to produce an audible signal with very little delay or system degradation. Because of the enormous cost savings and flexibility offered by VOIP, many technology analysts predict that by the year 2010, virtually all phone calls will be carried using VOIP technology. This same technology is also expected to be used for transmitting movies and other forms of entertainment.

Where to Learn More

Books

Seiter, Charles. *Everyday Math for Dummies*. Indianapolis: Wiley Publishing, 1995.

Web sites

A Brief History of Mechanical Calculators. "Leibniz Stepped Drum." <http://www.xnumber.com/xnumber/mechanical1.htm> (April 5, 2005).

History of Electronics. "Calculators." <http://www.etedeschi.ndirect.co.uk/museum/concise.history.htm> (April 6, 2005).

Homerun Web. "How to Calculate Earned Run Average (ERA)." <http://www.homerunweb.com/era.html> (April 12, 2005).

Mathworld. "Division." <http://mathworld.wolfram.com/division.html> (April 16, 2005).

Mathworld. "Trial Division." <http://mathworld.wolfram.com/trialdivision.html> (April 16, 2005).

The Math Forum at Drexel. "Dividing by 0." <http://mathforum.org/dr.math/faq/faq.divideby0.html> (April 16, 2005).

The Math Forum at Drexel. "Galley or Scratch Method of Division" <http://mathforum.org/library/drmath/view/61872.html> (April 16, 2005).

The Math Forum at Drexel. "Egyptian Division" <http://mathforum.org/library/drmath/view/57574.html> (April 16, 2005).

Domain and Range

Overview

A function is generally defined as a rule that partners to each number x in a set a unique number y in another set. The set of x values to which the rule applies is the function's domain, and the set of y values to which it applies is its range.

Calculus allows us to mathematically study rates of change and motion, something not precisely possible prior to discovery of the fundamental theorem of calculus. The domain and range of a function are the essence or foundation of algebraic equations and calculus formulas. Everyday uses include graphs, charts and maps.

Fundamental Mathematical Concepts and Terms

A function is a set of ordered pairs (x, y) such that for each first element x, there always corresponds one and only one element y. The domain is the set of the first elements and the range is the term given to name the set of the second elements. Often the domain is referred to as the independent variable and the range as the dependent variable.

The domain is the first group or set of values being fed or input into a function and these values will serve as the x-axis of a graph or chart.

The range is the second group or set of values being fed or input into a function with these values serving as the y-axis of a graph or chart.

A Brief History of Discovery and Development

The word function was first used mathematically by German philosopher and mathematician Gottfried Leibnitz (1646–1716) during his development of curve relationships. He used the term to describe a quantity relative to a curve such as a particular point of a curve or said curve's slope. Today the specific type of functions Leibniz referred to are called differentiable functions.

During the eighteenth century, Swiss mathematician and physicist Leonhard Euler (1707–1783) began using the word function to describe a formula involving various parameters. Over the next century, calculus and functions were being expanded upon and developed by German mathematician Karl Weierstrauss (1815–1897) who promoted developing calculus based upon arithmetic or number theory rather than geometry. By the end of the nineteenth century, mathematicians began defining mathematics using set theory thus seeking to describe

Key Terms

Dependent variable: What is being modeled; the output.

Function: A mathematical relationship between two sets of real numbers. These sets of numbers are related to each other by a rule that assigns each value from one set to exactly one value in the other set. The standard notation for a function $y = f(x)$, developed in the eighteenth century, is read "y equals f of x." Other representations of functions include graphs and tables. Functions are classified by the types of rules which govern their relationships.

Independent variable: Data used to develop a model, the input.

every mathematical object as a set. It was German mathematician Johann Peter Gustav Lejeune Dirichet (1805–1859) and Russian born mathematician Nikolai Ivanovich Lobachevsky (1792–1856) who almost simultaneously gave a formal definition of function. They defined a function as a special case of a relation, with a relation being described by such concepts as "is greater than" or "is equal to" in arithmetic.

Real-life Applications

COMPUTER CONTROL AND COORDINATION

All modern applications and technology use functions to determine the domain and range of a given problem. Every time you observe a graph, use a calculator, turn on a computer, drive an automobile, or even watch television, you are interacting with calculus and the concepts of domain and range. An example of this would be the computer components found in modern aircraft. Equations, formulas, and functions are all utilized and working in onboard computer systems to increase the safety of modern aviation. These computer systems help compensate for the instability of the aircraft, weight vs. wingspan length disparity, and a host of other variables related to flight.

CALCULATING ODDS AND OUTCOMES

Another example of common uses would be calculating risk or determining odds. Share values can be depicted as functions with domain and range and analysis of such data differentiate a wise from a foolish investment. Insurance companies use calculus formulas and functions to determine the risk associated with insuring. Analysts and researchers use set theory and ratios as a method of evaluation.

Having a "working understanding" of functions—and of the components of domain and range (especially which are dependent and independent variables) allows deeper understanding of graphs, charts, maps, finance, and business strategies.

PHYSICS

Physics relies heavily on calculus. English physicist and mathematician Sir Isaac Newton (1642–1727), who independently developed calculus about the same time as did Liebniz, used the concepts of domain and range of a function in advancing a Law of Gravity. In 2005, physicists use the domain and range of functions in designing and solving equations relative to nuclear energy development, nuclear fission processes, and other scientific experiments. Quantum mechanics, a rudimentary physical theory which refers to discrete units that the theory assigns to certain physical quantities, is the underlying structure in many fields of chemistry and physics. The fundamental beginnings of quantum mechanics stem from functional analysis.

ASTRONOMERS

Astronomers use the domain and range of functions to plot trajectories, calculate distances, measure space and speed of objects, and more. One example would be calculating the future path of asteroids, comets, and falling space debris. A complex set of functions are used to determine where to search for stars, planets, black holes, comets, asteroids, and other objects in our own galaxy as well as in other galaxies. Data sorted into tables by domain and range helped prove the existence of the planets Neptune and Pluto.

ENGINEERING

Determining the design and construction of highways, roads, wiring layouts, inventing new molecules, or developing molecular structures all require the use of domain and range in order to achieve the necessary

Figure 1. A one year chart of stock prices. (Share price is represented in U.S. dollars on the y axis.)

results. Structural engineers use functions to determine tensile strength of metals, to study aerodynamics of vehicles, and in the design of the subtle aerodynamic curves of an automobile.

COMPUTER SCIENCE

Computer hardware and software all use domain and range in the design and implementation of their programs. Computer programming languages such as Functional programming, is structured upon the use of functions. This language is a programming paradigm that relates computation as the evaluation of mathematical functions. Inductive logic programming also uses functions in a declarative programming paradigm that is concerned with finding general rules based on a sample of facts.

GRAPHS, CHARTS, MAPS

The ability to interpret graphs, charts and maps requires the understanding of the domain and range of functions. The *x* axis of a graph is the domain or all input elements of a function, and the *y* axis is all the actual elements derived from the function. The same is true for charts; the horizontal line represents the larger set, and the vertical line represents the specific information derived from the larger set. (See Figure 1, a typical chart of stock prices.)

Satellite technology, space exploration, computer programming and languages, automobile design, nuclear physics and countless other scientific fields all rely on the interrelation of domain and range.

In this digital information age, privacy concerns are of increasing importance. Cryptography is a field which uses the domain and range of functions in order to develop keys and ciphers to hide important or sensitive information transmitted over the Internet or other media. For example, based upon a random number, the independent and dependent variables can construct an elaborate system of codes and keys.

Where to Learn More

Books

Krantz, Stephen G. *Calculus Demystified*. New York, NY: McGraw-Hill, 2003.

Websites

Husch, Lawrence S. Visual Calculus. 2001 <http://archives.math.utk.edu/visualcalculus/> (March 1, 2005).

Elliptic Functions

Overview

Elliptic equations are a type of mathematical expression that is related to the geometric shape called an ellipse. An ellipse is a shape like a flattened circle. To draw an ellipse, one picks two points (called "foci") and encloses them with a curve drawn so that the sum of the distances from the foci to every point on the curve is always the same. An ellipse is flatter or more stretched if its foci are farther apart.

Fundamental Mathematical Concepts and Terms

An ellipse can also be defined using an equation. However, the equation that defines an ellipse is not what mathematicians mean when they speak of an "elliptic equation." An elliptic equation is a particular type of equation that arises from calculating the length of part of an ellipse. The lengths of curves are calculated using the mathematical technique called "integration." Integration of part of an ellipse produces a type of function (a function is a rule for going from one set of numbers to another) called an "elliptic integral." An elliptic equation is the opposite or inverse of an elliptic integral, the "inverse" of a function being a second function that undoes the first one. Elliptic equations are thus related to ellipses, but in a rather roundabout way.

Technically speaking, an elliptic function is a function in the complex plane that is periodic in two directions. This statement needs some explanation. The "complex plane" is a flat space like the *x*-*y* plane that is used in ordinary geometry and in graphing, except that one of the directions or axes is reserved for so-called "imaginary" numbers, which have special properties. A function is "periodic" if it repeats itself after some distance. A zigzag line is an example of a function that is periodic in one direction; the rings of a bull's-eye target are an example of a function that is periodic in all directions.

Elliptic functions are an important tool for mathematicians and physicists, because they crop up during the solution of many larger, more complex problems. For example, they appear in the exact mathematical description of pendulums and tumbling objects.

Real-life Applications

CONFORMAL MAPS

Flat maps have useful properties: they can be shown on computer screens and take up little room when printed on paper. However, most of the things that we are

Key Terms

Derivative: The limiting value of the ratio expressing a change in a particular function that corresponds to a change in its independent variable. Also, the instantaneous rate of change or the the slope of the line tangent to a graph of a function at a given point.

Integral: A quantity expressible in terms of integers. Also, a quantity representing a limiting process in which the domain of a function is divided into small units.

interested in making maps of—like the surface of the Earth—are curved. And whenever a curved surface is mapped to a flat surface, there is distortion or error. The Mercator projection of the Earth's surface (any flat map of a curved surface is called a "projection") is the most widely used flat map of the world, and it distorts the Earth by grossly expanding shapes near the North and South Poles. However, the Mercator projection does have one useful property: it is "conformal." Elliptic functions are involved in the mathematics of conformal projections. A conformal projection does not squeeze or stretch small shapes. Thus, a small circle or triangle on the globe is still a circle or triangle on the Mercator map. Conformal projections are used to map the surfaces of the Earth and of other planets. They are also used by researchers who want to make flat maps of the complexly folded surface of the brain, and are being studied by plastic surgeons as a way to describe and predict the effects of surgery on the nose.

E-MONEY

Many companies in Europe, Japan, and the United States are developing forms of e-money, also called digital cash. E-cash is electronic money on a card. A certain amount of money value is programmed into a memory circuit in the card, the card is swiped through a machine at the store when you buy something, and the cost of the purchase is subtracted from the value on the card. Unlike a debit card, you don't have to have a bank account to use an e-money card. Value is loaded directly into the card.

To keep people from simply programming more money into their e-money cards and becoming instant millionaires, a secret code or "cryptosystem" is used to check that e-money is real when it is spent. The kind of cryptosystem proposed for e-money is "public-key cryptography," and the particular kind of public-key cryptography that is proposed is elliptic-curve cryptography. Elliptic-curve cryptography doesn't require long, complicated calculations, which makes it ideal for use with the relatively simple circuits that can be embedded in e-money cards. Elliptic functions are involved in the mathematics of elliptic-curve cryptography.

THE AGE OF THE UNIVERSE

In 1915, the German physicist Albert Einstein (1879–1955) published his theory of general relativity. According to this theory (which has been checked against experiment many times, and is used every day in the global positioning system [GPS]), space does not go on forever. In a way that cannot be pictured in the mind but can be described mathematically, it bends around on itself. Space is finite or limited in size. Furthermore, it is expanding—and if it is finite and expanding, there must have been a time when it began expanding from zero size. In other words, the Universe must have had a beginning. Calculating the age of the Universe from the theory of general relativity involves the use of elliptic functions. Combining such calculations with astronomical observations has shown that the Universe is approximately 14.5 billion years old.

Where to Learn More

Books

Straffin, Philip, ed. *Applications of Calculus*, Vol. 3. Washington, DC: Mathematical Association of America, 1997.

Web sites

Wolfam. "Elliptic function." <http://mathworld.wolfram.com/EllipticFunction.html> (Feb 13, 2005).

Estimation

Overview

Estimation is the act of approximating an attribute such as value, size, or amount. Estimation has applications in all walks of life—from how much salt to put on popcorn, to using complex mathematical methods to predict the economy. Many common ideas, such as estimating the distance from Earth to the Sun, would be inconceivable without mathematical estimation.

Fundamental Mathematical Concepts and Terms

Estimation is an essential tool in many mathematical situations. For example, many people have to set an alarm clock to make sure that they wake up in time to get ready and get to school. The first time they set their alarm, they probably had to think about the different tasks they needed to accomplish in the morning in order to get to school on time. Those tasks may be broken down in the following manner:

- School starts at 8:00 a.m.
- It takes about 20 minutes to walk to school.
- It takes about 30 minutes to make and eat breakfast.
- It takes 20 minutes to shower, brush their teeth, and get dressed.
- They may plan to press the snooze bar twice, so they set the alarm for about 20 minutes before they actually need to get up.

Adding all these amounts of time together gives an estimation that it takes about an hour-and-a-half to get ready and get to the destination. According to this estimation, these people need to set the alarm for about 6:30 a.m to be able to arrive on time. The ability to make a good estimate considerably depends on known information and past experience. Final estimations also often depend on previously established estimations. For instance, the amounts of time that they spend performing the various tasks are estimates in themselves. In this case, the final estimation is based on:

- Information provided—classes begin at 8:00 in the morning.
- Past experience—people have showered and brushed their teeth before; they have made and eaten breakfast before; they know if they are snoozers and that they usually push the button twice before they get out of bed.
- Rough estimation—they know how far school is from their house and about how long it takes to walk that distance; because they prefer to get to school early rather than late, so they overestimate the time just a bit.

Estimates are usually refined as they are tested. Each time a person wakes up, gets ready, and walks to school, they may make adjustments to their estimate. For instance, sometimes they may want to get to school earlier, so they set their alarm even earlier for that day. On the other hand, they may have the first class of the day canceled, and they decide to set the alarm for later so that they can sleep a little longer.

Defining acceptable levels of error is another key concept in making meaningful estimates. There are many different theories and methods for analysis of error in estimation.

Throughout history, multiple methods of estimation have been proposed and scrutinized. Controversy exists among the many estimators and analysts about which methods yield the most accurate results.

A Brief History of Discovery and Development

The word estimate is derived from a late sixteenth century Latin word meaning to determine, appraise, or value. However, various methods of estimation have been used throughout history.

An early account of mathematical estimation involves a question posed by Greek mathematician Archimedes (born c. 287 B.C.), in which he contemplated how many grains of sand would be needed to match the volume of Earth. In ancient times, the issue of understanding and labeling very large (and small) numbers posed a serious problem that hindered the capabilities of mathematicians. This issue is at least partly attributed to the limitations of the existing numbering system, which was much like the Roman numeral system. By utilizing his own numbering system (similar to the exponential numbering system adopted later), Archimedes was able to grasp numbers large enough to approximate key values for determining a reasonable estimation of the amount of sand required to fill the volume of the planet. His new notation also allowed him to convey his ideas to his peers and to effectively convince them of the relative accuracy of his estimate.

Archimedes employed existing geometric theory (the equation for finding the volume of a sphere), a commonly accepted estimation (an approximate value of the radius of Earth, the distance from the center to the crust), and an observed measurement (he estimated that an average grain of sand was basically spherical and had a diameter of about 1/100th of an inch). Using these tools, Archimedes was able to estimate that it would take over 10^{32} grains of sand to match Earth's volume. He was aware of many imperfections in his calculations, including the fact that every grain of sand is not perfectly spherical. Earth is not a perfect sphere either—it has mountains and canyons (and is squashed at the poles). The values for the radius of Earth and the average diameter for a grain of sand are also only approximate values. Nevertheless, he was able to derive an estimate that was substantial enough to give a manageable account of the magnitude of the answer to his question. This was an important step toward mankind becoming comfortable with previously unfathomable numbers.

In 1773, Benjamin Franklin found that a drop of oil placed on the surface of water will spread out across the water until it forms a layer that has the thickness of a single molecule of the oil (known as a monomolecular film). In addition to Franklin's immediate observations from this experiment (his notes describe the oil spreading quickly and causing wavy water to become calm almost immediately), his discoveries would have a profound influence in the first estimation of the thickness of a molecule more than a century later. Scientists first estimated the thickness of a molecule by recording the volume of a drop of oil and then placing the drop onto the surface of some water. Once the oil had spread to a thickness of a single molecule, the surface area of the floating film was estimated using an approximate value for the radius of the somewhat circular layer of film. Then the formula for volume was used to determine the thickness of the film. The volume of the oil is equal to the area of the film multiplied by the thickness of the film. So to determine the thickness of the film, which is the thickness of a single molecule, the volume of the drop of oil was divided by the approximate area of the film.

In 1801, Carl Frederick Gauss, also commonly viewed as one of the most important figures in the history of mathematics, made the first applicable estimation of the orbit of planets. His first subject was a newly discovered planet named Ceres. Using his method of least square, which remains an important contribution to the development of estimation methods, Gauss was able to enhance prior theories about the orbits of planets, incorporating calculations that represent imperfections in orbital paths due to factors such as interference caused by other celestial bodies.

From tracking and maintaining populations of endangered species, to predicting genetic abnormalities in unborn babies, estimation remains an essential concept in many mathematical and scientific procedures and discoveries.

Real-life Applications

BUYING A USED CAR

For most people, the purchase of their first car can be an overwhelming event. There is much more to consider than whether they like the color or the rims on the

wheels. With so many factors that can affect the value of an automobile, people have to be careful that they don't get cheated.

Often, the seller will begin by asking for much more money than the car is actually worth, or keep any problems a secret until they officially sell the car. So, how can one know if they are getting a good deal or being cheated? First, one should be aware that issues might exist, such as engine trouble, damaged upholstery, or body damage that lower the true value of the car. Similarly, any enhancements that can raise the value must be considered, including custom parts, limited edition features, audio or video equipment, global positioning system (GPS) tracking devices, safety features, and so on.

The most significant obstacle in coming to an agreement between the seller and the buyer is that there is no true value of a used car. Too many factors influence each car to be defined generally. Luckily, one thing that the buyer and the seller will probably agree on is that they both want to finalize the transaction as quickly and smoothly as possible. Therefore, both people have to use estimation involving available information to find an agreeable price range. Most buyers and sellers alike refer to one of many periodically released publications, such as the *Kelley Blue Book*, as a basis for what the car should cost depending on the year, make, model, mileage, and general condition. Using this base price, both parties attempt to factor in as many positive and negative characteristics as they can find to determine what they think the car is worth and to arrive at lower and upper bounds for an acceptable price. From there, everything depends on keen negotiation skills.

GUMBALL CONTEST

Jen has been entered into the critical-thinking contest at the annual mathematics fair, in which the top math students from around the region compete to solve difficult problems. The first problem posed to the contestants involves the estimation of the number of gumballs contained in a glass case that is 4 feet long, 4 feet wide, and 8 feet tall. Each contestant is expected to use mathematical reasoning to decide whether they think that the number of gumballs inside the glass case is less than or greater than 25,000.

Jen examines the glass case and thinks about how she can make a good approximation of the number of gumballs inside. The first thing she does is collect as much information as she can about the problem at hand. Jen takes note of the following information:

- The glass case is transparent, so Jen can approximate the size of the gumballs. As far as she can see, the gumballs are all about the same size—somewhere between 1 inch and 2 inches in diameter.
- The volume of the glass case is equal to the product of its dimensions. Since she will be estimating the volume of the gumballs in cubic inches, Jen multiplies each dimension of the glass case by 12 to convert to inches. The glass case is 48 inches by 48 inches by 96 inches. Multiplying these values together, she finds that the volume of the glass case is 221,184 cubic inches.

Her estimate of the diameter of each gumball consists of an upper bound and a lower bound. In this way, she hopes to simplify the problem without concerning herself too much with the true size of one gumball—let alone all of them! If the total estimated volume that she finds using the lower bound for the size of a gumball is more than the volume of the case, then she will know that 25,000 gumballs will not fit into the case. If the estimate she finds using the upper bound for the size of a gumball is less than the volume of the case, then she can safely conclude that 25,000 gumballs will fit into the glass case. However, if the volume of the glass case is between her lower and upper estimates, then she cannot make a confident conclusion using this information, and she will have to attempt to refine her estimate of the size of a gumball.

Jen knows that she needs the size of the gumballs to be expressed in terms of volume so that she can compare the volume of 25,000 gumballs to the volume of the glass case. She needs to find the volume of a single gumball and then multiply this value by 25,000. Using the formula for the volume of sphere, she finds that a gumball that is 1 inch in diameter (having a half-inch radius) has a volume of approximately 2.09 cubic inches. Similarly, she finds that a gumball that is 2 inches in diameter (having a one-inch radius) has a volume of 4.19 cubic inches. At this point, she feels confident that the volume of a single gumball is somewhere between these two estimated bounds. To find bounds for the total estimated volume of 25,000 gumballs, she multiplies the bounds for the volume of a single gumball by 25,000. She is reasonably certain that the volume of 25,000 gumballs is between 52,359 and 104,750 cubic inches. Since the upper bound for her estimate of the total volume of the gumballs is less than the volume of the glass case, Jen could decide at this point that she is convinced that 25,000 gumballs will in fact fit into the case. However, just before she turns her response in for evaluation, she becomes aware of a major flaw in her reasoning.

The total estimated volume of the gumballs definitely gives her a better feeling for this problem, but she quickly realizes that this will not yield conclusive results

because she has not considered the air space in between all of the gumballs. What she has really figured out is that if she were to chew up 25,000 gumballs and press them into the glass case, the huge blob of gum would fit (especially if the gumballs are hollow). Little does Jen know that this is an example of a sphere-packing problem—a classic problem in mathematics for which there is no standard solution. Nevertheless, she is on the right track to making a fairly good estimate.

While thinking of a way to refine her estimation methods, she imagines having each gumball wrapped perfectly into a little box. She understands that this idea will lead to an overestimate of the amount of airspace in the glass case because the gumballs do not stack on top of each other perfectly. She will use the values found using this method as upper bounds for the total volume of the gumballs taking airspace into account.

Each gumball-wrapping box would have length, width, and height equal to the diameter of the gumball. For a gumball with a 1-inch diameter, the surrounding box would have a volume of 1 cubic inch. For a gumball with a 2-inch diameter, the box would have a volume of 8 cubic inches. The question now is whether or not 25,000 of these surrounding boxes will fit into the glass case. Multiplying by 25,000, she finds lower and upper bounds for the total volume of all of the wrapping boxes. The total volume of the boxes is between 25,000 and 200,000 cubic inches.

At this point, Jen stops to think about her progress so far. Since she can show that her largest estimate of 200,000 cubic inches—found by packing gumballs into boxes that overcompensate for airspace—is smaller than the volume of the glass case (221,184 cubic inches), she feels fairly certain that 25,000 gumballs fit into the glass case. (Note that if any of Jen's estimates were greater than 221,184 cubic inches, then she would have had to either refine her estimate of the diameter of a gumball or come up with a more accurate way to account for air space.)

POPULATION SAMPLING

Wildlife conservationists are often confronted with the task of estimating how many members of a certain species of animal are living in a given area. For example, suppose that a team of conservationists needs to estimate the number of fish in a small lake (without draining the lake and counting all of the fish). This may seem a daunting task because fish move around the lake, reproduce, and die. However, the team will be able to use population sampling techniques to find an estimate that is suitably accurate for their needs.

To begin, the team catches a sample of 300 fish. Each of these fish is tagged and returned to the lake. The team then makes a simplifying assumption that will be critical to the estimation process: over time, all of the fish in the lake move about at random. This is a reasonable assumption based on previous studies about these fish.

After waiting a week for the fish to redistribute themselves, the team again catches 300 fish and finds that 25 fish out of this sample are tagged. This time, the team members must do their best to select the fish at random from the total population of the lake. To ensure that they collect a random sample, they may collect the sample from various areas of the lake.

Next, the team uses the basic sampling principle as it applies to their situation: the proportion of tagged fish in the second sample should reflect the proportion of tagged fish in the entire lake population, as long as the sample size is reasonably large.

In the second sample, the team found that 25 out of 300 fish were tagged, so the proportion of tagged fish in the second sample is 25 divided by 300, or 1/12. The team also knows that there were 300 tagged fish in the lake (barring any fatalities among the first sample), so the proportion of tagged fish in the entire lake is 300 out of the total population of fish, the value that the team is attempting to estimate. In accordance with the basic sampling principle, the formula $1/12 \approx 300/N$, where N is the total number of fish in the lake helps team members find an estimate for the total population of fish in the lake.

After dividing both sides by 300 and simplifying, the team finds that $N \approx 3{,}600$. The team of conservationists now has a rough estimate of 3,600 fish living in the lake. Depending on the requirements of the study, the team may or may not need to take more samples and find an average value. The team would not replace the fish after each sample, so that the fish are not counted twice. A relatively consistent number of tagged fish in each sample would be a good indication that the estimations are sufficiently accurate. Using a larger sample will usually result in higher accuracy as well.

DIGITAL IMAGING

A digital image is an arrangement of tiny square regions called pixels. In the case of a gray-scale (black-and-white with shades of gray) image, the brightness of each pixel is determined by a numeric value. A typical gray-scale image contains values ranging from 0 to 255, with 0 representing black, 255 representing white, and intermediate values representing shades of gray.

A color image can be represented using different mixtures of red, green, and blue. The color of each pixel

in the digital image is usually determined by a set of three numbers, one representing red, one representing green, and one representing blue. These values each range from 0 to 255, where 0 indicates that none of that color is present in that pixel and 255 indicates a maximum amount of that color. When a digital image is magnified many times, the pixels can be seen clearly. If only part of a magnified image is visible, it may look like nothing more than different colored squares.

Estimation is a key concept in digital image compression processes. The goal of digital image compression is to reduce the size of the image file (so that it can be efficiently stored and shared) without losing so much quality that the human eye will easily notice the change. The main difference between image formats is the way that they compress images. The graphics interchange format (GIF) and joint photographic experts group (JPEG or JPG) formats—two of the most common digital image formats—are good examples of the effects of the various image compression techniques.

GIF images only support 256 colors—not much compared to the millions of colors found in most color photographs. If an image is converted to the GIF format, a compression technique called dithering is used to compensate for any loss of color. Image dithering involves repeating a pattern of two or more available colors in order to trick the eye into seeing a color very close to the color found in the original photograph. For example, to represent a solid area of a shade of red that is not included in the available 256 colors, the dithering process may alternate every other pixel with the two closest available shades of red. As the image is magnified, the pattern becomes more and more apparent. The color patterns that result from the dithering process are determined by mathematical functions that perform operations for estimating unavailable colors. The GIF format is best suited for illustrations and graphics with large regions of solid color. On the other hand, when a photograph containing millions of colors is converted to GIF, it usually appears grainy because there are too many colors to be adequately represented by just 256 colors.

JPEG images are much better suited for photographs. The JPEG format supports millions of colors and its compression method is intended to handle quick changes in color from pixel to pixel. However, graphics with relatively large areas of solid color converted to JPEG images tend to display messy spots around the areas where colors change. For example, if a company logo consisting only of a blue word on a solid red background is saved as a JPEG image, it will most likely have fuzzy areas all around the border of the letters. These fuzzy areas are called compression artifacts and, as implied by their name, are results of the compression process. As with the dithering process of the GIF format, the compression process is heavily dependent on mathematical functions that attempt to reduce the file size while retaining the image as seen through the human eye.

CARBON DATING

One of the most influential concepts in the field of archeology is that of carbon-14 dating, which allows archeologists to estimate the age of fossils and human artifacts. The basic idea behind carbon-14 dating is that all living things, from plants to humans, contain the same ratio of carbon-14 and carbon-12 atoms at all times (for every carbon-14 atom, there are a certain number of carbon-12 atoms). In a living organism, both types of atoms are constantly being created and destroyed, but the ratio between the two remains constant.

As soon as a living organism dies, it stops producing new carbon atoms. The carbon-14 decays and is no longer replaced, while the carbon-12 does not decay at all. By comparing the ratio of carbon-14 to carbon-12 in a formerly living organism to that in a living organism, it is possible to estimate how long the former has been dead. This concept has allowed archeologists to uncover many important milestones in the history of humankind.

Potential Applications

THE HUBBLE SPACE TELESCOPE

The Hubble Space Telescope (HST), a high-powered telescope attached to a spacecraft, has revolutionized astronomy by allowing astronomers to view celestial sights that are billions of light years away. Due to the fact that the images are captured from billions of light years away, astronomers know that the events depicted took place billions of years ago! Breathtaking images that have been constructed using data transmitted from the HST can be found on the Internet, in books, magazines, and newspapers.

However, these images are not exact representations of what is truly out there. The HST is capable of detecting different types of light and heat, including visible light (that humans can see), ultraviolet light, and infrared light. The raw data transmitted by the HST are electronic black-and-white images that reveal very little detail. Astronomers must combine the data from the various images (created from the different types of light and heat) and interpret the overall picture. These interpretations require advanced estimation methods, as well as some

imagination and creativity. Since its launch in 1990, the HST has undergone many revisions, including updates to its image-capturing tools. As space telescope technology is refined, astronomers are able to construct increasingly accurate representations of celestial activity, providing valuable insight into the vastness of the universe.

SOFTWARE DEVELOPMENT

Software developers strive to estimate the amount of time that it will take to complete a software development cycle. A single development cycle often involves a vast number of steps that can take anywhere from a few weeks to a few years. All of these steps must be accounted for in the development plan to ensure that the software is completed, tested, and revised in a specified amount of time. If the product is not ready on time, the software company may lose clients and funding. Some of the major steps in the development cycle include conception (coming up with initial ideas), planning (organizing ideas and time), design (working out the look and feel of the on-screen display and general functionality), coding (using a programming language to write the software), and testing (checking to make sure that things look and work correctly). All of these steps are made up of multiple smaller tasks. For example, design might be split into visual design and functional design. Visual design might be split into window design, menu design, and so on. If any part of testing fails, issues must be listed and categorized by severity. The development team must then go back to the development cycle. How far back in the cycle the development team must go depends on the issues found by the testing team.

For decades, people have tried to conceive a universally accepted method for estimating the time it will take to complete a software product. However, software companies continue to run into unforeseen snags in the process, causing them to miss deadlines. Compensating for less tangible aspects of the development process proves to be a difficult task. For example, the complexity of the project (the number of pages, the types of tasks the software performs, how information is processed, etc.) is a consistent source of error.

In spite of past limitations to software development strategies, the desire to streamline the development processes continues to grow. This is due to a steady increase in the demand for software products, a trend that is not expected to change in the near future.

Where to Learn More

Books

Klette, R., and A. Rosenfeld. *Digital Geometry—Geometric Methods for Digital Picture Analysis.* San Francisco: Morgan Kaufmann, 2004.

Periodicals

Lewis, A. P. "Large Limits to Software Estimation." *ACM Software Engineering Notes.* 26, No. 4 (2001): 54–59.

Web sites

Calkins, Keith G. "The How and Why of Statistical Sampling." Andrews University. October 4, 2004. <http://www.andrews.edu/~calkins/math/webtexts/stattoc.htm> (March 8, 2005).

U.S. Census Bureau. "About Population Projections." August 2, 2002. <http://www.census.gov/population/www/projections/aboutproj.html> (March 9, 2005).

Exponents

Overview

An exponent is a number placed just above and to the right of another number to say how many times the lower number should be multiplied by itself. For example, $2^3 = 2 \times 2 \times 2$, where the exponent is 3.

We can handle some very large and very small numbers easily using exponents. For example, instead of 100,000,000,000,000,000,000 can write 10^{20}. Equations that contain a variable as an exponent, such as $y = 5^x$, are known as exponential equations. They are used to describe the breakdown of radioactive atoms, the growth of living populations, the interest paid on loans, the cooling of planets and other objects, the spreading of epidemic diseases, and many other situations.

Fundamental Mathematical Concepts and Terms

BASES AND EXPONENTS

The expression 2^3 is read as "two to the third power" or "two to the power of three." Here 3 is the exponent and 2 is the "base." In 5^6, the exponent is 6 and the base is 5.

INTEGER EXPONENTS

An integer is a whole number, like 3, 0, or −12. When a positive integer like 3 is used an exponent, it tells us to take the base and multiply it by itself. Thus, for example, $10^4 = 10 \times 10 \times 10 \times 10 = 10{,}000$. (Notice that when 10 is the base, the exponent gives the number of zeroes in the product.) For any number a and any positive integer n,

$$a^n = \overbrace{a \times a \times a \cdots a}^{n \text{ times}}$$

For any two positive integers, which we can call m and n, $a^m a^n = a^{m+n}$. For example, if the base is $a = 10$ and the exponents are $m = 2$ and $n = 3$, then

$$\begin{aligned} 10^2 10^3 &= \overbrace{10 \times 10}^{2 \text{ times}} \times \overbrace{10 \times 10 \times 10}^{3 \text{ times}} \\ &= \overbrace{10 \times 10 \times 10 \times 10 \times 10}^{2+3 \text{ times}} \\ &= 10^{2+3} = 10^5 \end{aligned}$$

Several other useful rules apply to integer exponents such as that $(a^m)^n = a^{an}$ or that $(ab)^m = (a)^m(b)^m$. Here are examples of these rules in action:

- $(a^m)^n = a^{mn}$ means that $(10^2)^3 = 10^2 10^2 10^2 = 10^{2+2+2} = 10^{2\times 3} = 10^6$
- $(ab)^m = a^m b^m$ means that $(3\times 10)^2 = 3^2 10^2$

As for negative integer exponents, they also have a simple meaning:

$$a^{-n} = \frac{1}{a^n} = \frac{1}{\underbrace{a\times a\times a\cdots a}_{n \text{ times}}}$$

What about using 0, which is neither positive nor negative, as an exponent? By definition, $a^0 = 1$ for any number a other than 0 itself. For example, $1^0 = 1$, $-10^0 = 1$, and $1{,}000{,}000^0 = 1$. But this doesn't work for 0^0. Raising 0 to the power of 0, like dividing by 0, is what mathematicians call "undefined"—it has no meaning. You might want to try raising 0 to the power of 0 (or dividing anything by 0) on your calculator, and see what happens.

NON-INTEGER EXPONENTS

So much for integer exponents. But how do we handle an expression with a fractional exponent, like $2^{1/3}$? We can't multiply 2 by itself one-third of a time! Therefore, we expand our definition of exponent to include rational numbers, that is, all numbers that can be written as fractions, such as 1/3. The rational numbers include the integers, because we can always write an integer as a fraction by putting a 1 in the denominator: 56 = 56 / 1. Any number in decimal form, such as 5.34, can also be written as a fraction:

$$5.34 = 5 + \frac{3}{10} + \frac{4}{100} = \frac{534}{100}$$

Let's start with rational numbers of the form

$$\frac{1}{n}$$

where is a positive integer. For two positive numbers a and b, $b = a^{1/n}$ means that $b^n = a$. For example, $3 = 9^{1/2}$ means that $3^2 = 9$, and $5 = 25^{1/2}$ means that $5^2 = 25$. When $b = a^{1/2}$, as in these two examples, we say that b is the "square root" of a; so 3 is the square root of 9, and 5 is the square root of 25. Taking the "square" of b (raising b to the power of 2) gives a back again: $3^2 = 9$ and $5^2 = 25$.

When $b = a^{1/3}$ we say that b is the "cube root" of a, meaning that $b \times b \times b = a$. When $b = a^{1/n}$ we say that b is the "nth root" of a, meaning that $b \times b \times b \ldots \times b$ (n times) $= a$.

By combining this rule for $1/n$ exponents with the rule that $a^{mn} = (a^m)^n$, we can see what it means to use rational numbers (fractions) as exponents, as in $a^{m/n}$: namely, $a^{m/n} = (a^m)^{1/n}$. And we already know how to deal with exponents like m and $1/n$ separately. For example,

$$\begin{aligned} 3^{3/2} &= (3^3)^{1/2} \\ &= (3\times 3\times 3)^{1/2} \\ &= 27^{1/2} \end{aligned}$$

$27^{1/2}$, the square root of 27, is approximately 5.1961524. To write it down exactly, we would have to write an infinitely long string of digits to the right of the decimal point.

We've been looking at the meaning of rational exponents—exponents that can be expressed as fractions with integers in their numerators and denominators. Any number that can't be represented as a ratio of integers, like π, is termed irrational. Since we can't express an irrational number as a fraction, our method for dealing with rational exponents won't work for irrational exponents. The irrational exponent must be approximated as a rational exponent before it can be evaluated.

EXPONENTIAL FUNCTIONS

A function is a rule that relates numbers to each other. For example, the function $f(x) = 2x$ ("f of x equals $2x$") means that for every number x there is another number, $f(x)$, that is related to it by being twice as large.

The exponential function is $f(x) = b^x$, where b is any number other than 1. The function behaves differently depending on whether x is greater than 1 or between 0 and 1. If b is greater than 1—say, $f(x) = 2^x$—then the exponential function behaves as shown in Figure 1.

Figure 1 shows the plot of the exponential function $f(x) = 2^x$. All functions of the form $f(x) = b^x$ with $b > 1$ have this shape, and all equal 1 at $x = 0$. The curve in this figure looks like it touches the x axis at the far left, but the curve never quite gets there, no matter how negative x becomes.

The key features of $f(x) = 2^x$ are its slow decline to the left, like a plane coming in for a landing that never quite touches the runway, and its upward zoom to the right. The curve increases to the right because we are raising 2 to increasingly large exponents : for $x = 2$ we have $f(2) = 2^2 = 4$, for $x = 5$ we have $f(5) = 2^5 = 32$, and so on. The functions tails off toward 0 as x gets more negative because we are raising 2 to increasingly negative exponents:

$$f(-1) = 2^{-1} = \frac{1}{2},\ f(-5) = 2^{-5} = \frac{1}{2^5} = \frac{1}{32}$$

and so on.

If b is between 0 and 1, the exponential function $f(x) = b^x$ behaves as shown in Figure 2.

All functions $f(x) = b^x$ with $0 < b < 1$ have a similar shape, and all equal 1 at $x = 0$.

What do we do with negative exponents in an exponential equation, $f(x) = b^{-x}$? This can be rewritten using the rule that $(a^m)^n = a^{mn}$ $f(x) = b^{-x} = b^{(-1)(x)} = (b^{-1})^x$. Since

$$b^{-1} = \frac{1}{b}$$

using a negative exponent is the same thing as using a positive exponent with the base flipped upside down:

$$b^{-x} = \left(\frac{1}{b}\right)^x$$

For example,

$$2^{-x} = \left(\frac{1}{2}\right)^x$$

The rule $(a^m)^n = a^{mn}$ also tells us how to deal with numbers that multiply the exponent, as in $f(x) = b^{ax}$. We can always rewrite the function so that the exponent is plain old x: $f(x) = b^{ax} = (b^a)^x$.

Table 1 presents a summary of the laws of exponents.

The concept of the exponent boils down to repeated multiplication: take a number b, multiply it by itself, multiply that result by b, multiply that result by b, and so on. People began to play with this concept—geometric progression, as it is also called—very early in history.

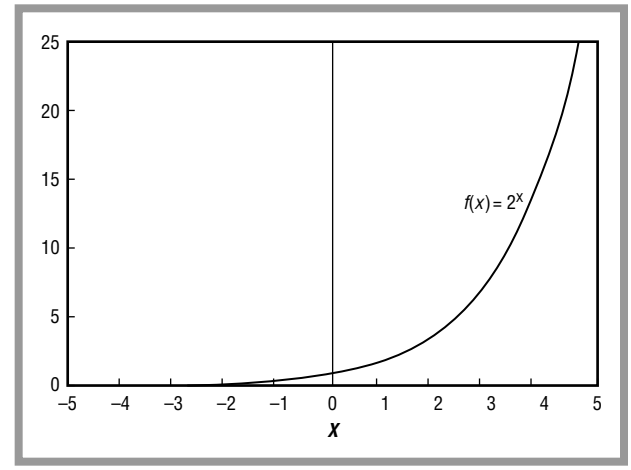


Figure 1: Plot of the exponential function $f(x) = 2^x$.

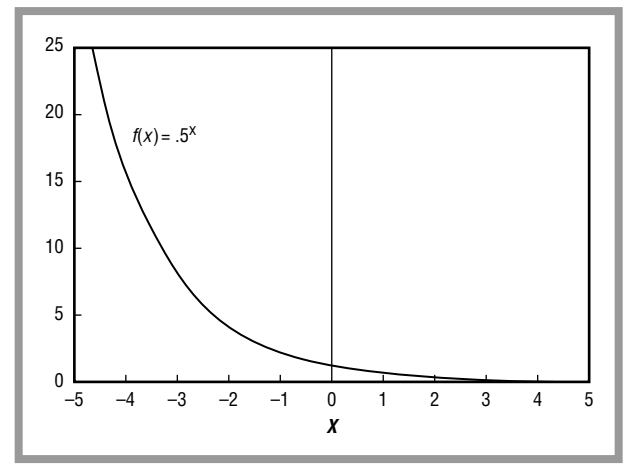


Figure 2: Plot of the exponential function $f(x) = (1/2)^x$.

Geometric progression was studied over 3,600 years ago by the ancient Egyptians and Sumerians and much later by the Greeks, including Euclid (c. 300 B.C.) and Archimedes (287?–212 B.C.).

A Brief History of Discovery and Development

Especially after the Middle Ages, scientists became aware of many real-life objects that behave in a geometric

Rule	Meaning	Example
$a^n a^m = a^{n+m}$	Multiplying two exponential terms having the same base is the same as raising that base to the product of the two exponents.	$2^2 2^3 = 2^5 = 32$
$\frac{a^n}{a^m} = a^{n-m}$	Dividing two exponential terms having the same base is the same as raising that base to the difference of the exponents.	$\frac{2^3}{2^2} = 2^{3-2} = 2^1 = 2$
$\left(a^m\right)^n = a^{mn}$	Applying an exponent to a base that is already raised to an exponent is the same as raising that base to the product of the two exponents.	$\left(2^3\right)^2 = 2^{3\times 2} = 2^6 = 64$
$\left(ab\right)^n = a^n b^n$	Raising the product of two numbers to an exponent is the same as raising each number separately to that exponent and then multiplying.	$(3\times 5)^2 = 3^2 5^2 = 9\times 25 = 225$
$\left(\frac{a}{b}\right)^n = \frac{a^n}{b^n}$	Raising a fraction to an exponent is the same as raising the numerator and denominator separately to that exponent and then dividing.	$\left(\frac{3}{5}\right)^2 = \frac{3^2}{5^2} = \frac{9}{25}$

Table 1: Summary of the laws of exponents.

or exponential way, including the unrestrained breeding of animals; the cooling of hot objects; the shapes of natural spirals such as those found in pine cones, sunflowers, and ram's horns; the dimming of supernovae (exploding stars); the relationships between musical notes; and many others. Ancient records inscribed in clay show that in the Middle East, the Sumerians knew about the exponential properties of compound interest as early as 2000 B.C.

Our modern way of writing an exponent-placing a small number above and to the right of another number—was introduced in 1637 by the French philosopher and mathematician René Descartes (1596–1650). At abut that time the relationship between the logarithm and the exponent (namely, that they are inverses of each other) was finally clarified.

In the eighteenth century, Swiss mathematician Leonhard Euler (1701–1783) first devised the complex exponential function, where a base is raised to the power of an "imaginary" number containing the square root of -1. The square root of -1 was a radical new idea because it seemed impossible: what number, when multiplied by itself, could give -1? The square root of $+1$ is simply itself (because $1^2 = 1$), but -1 cannot be its own square root (because $-1^2 = -1 \times -1 = +1$). Nevertheless,

mathematicians have found numbers containing the square root of -1, called "complex" numbers, to be very useful. Euler also explored the use of the number $e = 2.7182818$ as a "natural" (that is, highly convenient) base for exponents and logarithms.

Today, exponents are applied throughout mathematics. They are used in physics and engineering to describe phenomena that fade with time, such as radioactivity, or periodic (repeating) phenomena like waves. They are used in biology to model populations of bacteria, animals, and people; in medicine to model the breakdown of drugs in the body; and in business and economics to describe interest and inflation.

Real-life Applications

SCIENTIFIC NOTATION

With scientific notation you can write down a number greater than the number of atoms in the universe with just a few strokes of your pen. This is how:

Recall that for powers of ten, the exponent gives the number of zeroes: $100 = 10^2$, $1000 = 10^3$, and so forth. Using this fact, an ugly number like 1,000,000,000,000,000,000 becomes a user-friendly 10^{18}. We can also write a number like 1,414,000,000,000,000,000 as 1.414 times 10^{18}, namely, 1.414×10^{18}. A number written in this form is said to be in "scientific notation." Scientific notation makes large numbers much easier to handle.

The same trick also works for small numbers, because numbers like .1, .01, .001, and so forth can be written as tens raised to negative exponents; for example,

$$.01 = \frac{1}{100} = \frac{1}{10^2} = 10^{-2}$$

We can therefore write 10^{-20} instead of .00000000000000000001, and 1.675×10^{-24} instead of .0000000000000000000001675 (which happens to be the mass in grams of a single hydrogen atom).

Another useful feature of scientific notation is that changing the exponent is shorthand for moving the decimal point. Thus we write 4.5×10^{-2} for .045 and 4.5×10^{-4} for .00045.

To multiply two numbers written in scientific notation, all we have to do is multiply the numbers out front and add the exponents. So, 1.414×10^{18} times 1.675×10^{-24} is $1.414 \times 1.675 \times 10^{18-24} = 2.36845 \times 10^{-6}$. This is so easy that when multiplying simple numbers like 1×10^{12} and 5×10^9, it's actually easier to do the math in your head than to punch buttons on a calculator provided you can add 12 and 9 in your head without blowing a fuse. (The answer is $(1 \times 10^{12}) \times (5 \times 10^9) = (1 \times 5) \times 10^{12+9} = 5 \times 10^{21}$.) Division is equally easy, only you divide the numbers out front and subtract the exponents.

As for writing down a number larger than the number of atoms in the universe, this is it: Physicists estimate that there are fewer than 10^{100} atoms in the Universe, perhaps only about 10^{80}. If every atom in the Universe were inflated into a universe full of atoms, there would still be only $(10^{80})^2$ or 10^{160} atoms in existence. So writing 10^{160}, or 10^{200}, or 10^{300} gives a number much, much greater than the number of atoms in the Universe.

EXPONENTIAL GROWTH

A useful fact in science (and banking) is this: Any quantity that grows by a fixed percentage during each interval of time grows exponentially.

Consider a pair of rabbits. Say that this pair has two offspring by the end of one year. There are now four rabbits: the population has doubled in one year. Assume that both pairs will breed the following year, each producing two more offspring, and that their offspring will also breed, and so on, so that the total population keeps on doubling every year. This is the same as saying that the population increases by a fixed percentage every year, namely 100%: 2 rabbits plus 100% of 2 rabbits equals 4 rabbits (first year's population growth), 4 rabbits plus 100% of 4 rabbits equals 8 rabbits (second year's growth), and so forth. The growth of this rabbit population is described by an exponential equation. If we label the first year Year 0, then the number of rabbits at the beginning of each year, $R(t)$. is given by the following series of numbers: $R(0) = 2$, $R(1) = 4$, $R(2) = 8$, $R(3) = 16$, and so on. In general, $2R(t) = 2 \times 2^t$, where t is years.

Figure 3 shows the number of rabbits as an exponential function of time, starting with two rabbits and assuming a doubling time of 1 year. This curve is similar to the part of Figure 1 to the right of $x = 0$.

Do rabbits really do this? Sure—when they can; that is, if they have food to eat, room to live in, and air to breathe, and no enemies or diseases nasty enough to keep the population from growing. In reality, rabbits cannot have all these things, certainly not in infinite amounts. There are predators and diseases; there is only so much food, water, and room. So in one sense the exponential equation is not realistic—but in another sense, it is terribly realistic. It forces conflict between growing populations and their environments. If a population seeks to increase by any fixed percentage per year (that is, if it seeks to grow exponentially), then at some point

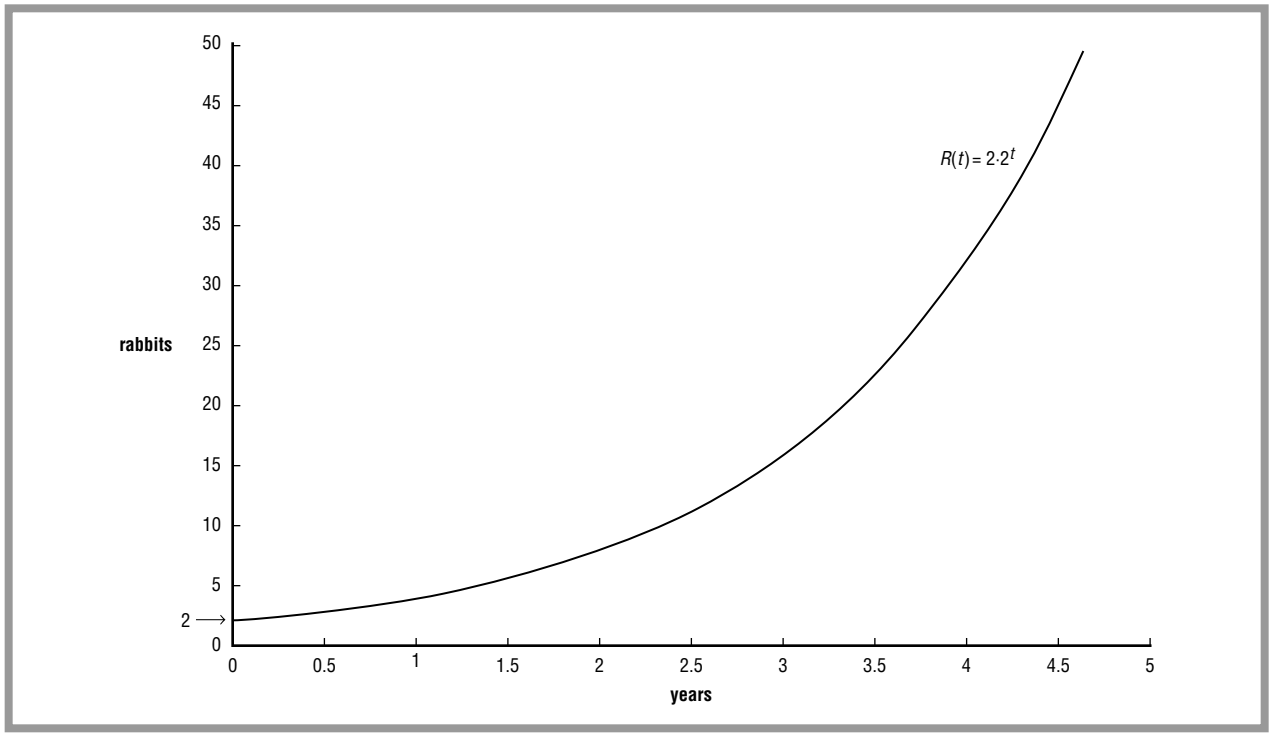


Figure 3: An exponential function.

deaths—whether from hunger or other causes—will inevitably outpace births. To see why this cannot be avoided, let's use the master rabbit equation, $R(t) = 2 \times 2^t$, to look at what would happen if our imaginary rabbit population was somehow, magically, able to keep on growing.

Let's calculate how long it takes to get a given number of rabbits, N. To do so, we find the "solution" to the equation $2 \times 2^t = N$, that is, that unique value of t for which the equation is true. Let's pick a nice, big value for N—say, enough rabbits to completely fill the Solar System. Pluto, usually regarded as the outermost planet, has an average distance from the Sun of 5.914×10^{12} km. Because the volume of a sphere of radius r is

$$\frac{4}{3}\pi r^3$$

(exponents again!), the volume of the Solar System is

$$\frac{4}{3}\pi(5.914\times10^{12})^3\,\mathrm{m}^3 = \frac{4}{3}\pi 5.914\times10^{36}\,\mathrm{m}^3$$

where 1 m^3 is a cubic meter (the amount of space in a cube 1 meter across). If we can pack 50 rabbits into each cubic meter of space, then the number of rabbits that can fit into the Solar System is

$$N = \overbrace{\left(\frac{4}{3}\pi 5.914\times10^{36}\,\mathrm{m}^3\right)}^{\text{number of cubic meters in Solar System}} \times \overbrace{\left(\frac{50\text{ rabbits}}{\mathrm{m}^3}\right)}^{\text{number of rabbits per cubic meter}}$$

$$= \frac{4}{3}\pi 2.957\times10^{38}\text{ rabbits to fill Solar System}$$

Because the number of rabbits at the start of year t is $N = 2 \times 2^t$, to find out the number of years till there are

$$N = \frac{4}{3}\pi 2.957\times10^{38}$$

rabbits we need to find t such that

$$N = \frac{4}{3}\pi 2.957\times10^{38}$$

To find the t that satisfies this equation, we must perform the mathematical operation known as "taking the logarithm" of both sides. Taking the logarithm undoes

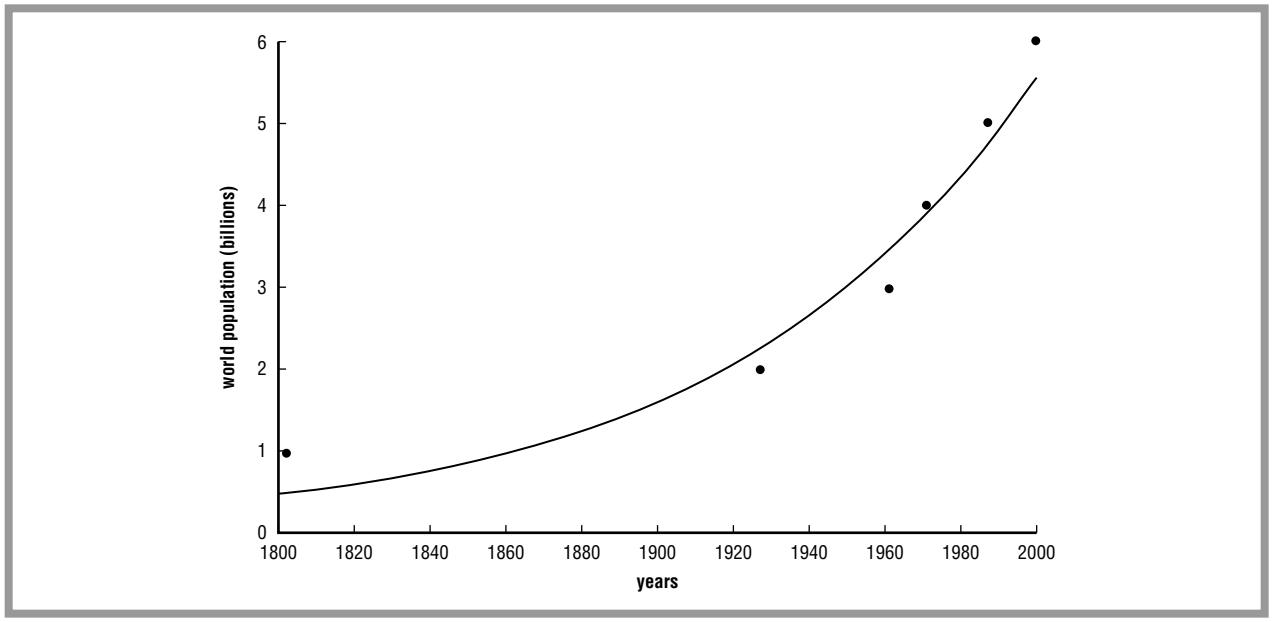


Figure 4: Population growth.

exponentiation (applying exponents) in much the same way that subtraction undoes addition or division undoes multiplication.

Solving using logarithms, we find that t equals approximately 129 years. This means that a rabbit population that doubled every year would fill the whole Solar System with long-eared rodents in only 129 years. It would take the first 128 of those years to fill half the Solar System, but just one more year to fill the other half!

A Solar System full of rabbits is, of course, physically impossible. The moral is that exponential population growth always runs up against physical limits, most often getting eaten or starving to death.

The equation $R(t) = 2 \times 2^t$ is an example of the general exponential equation $R(t) = R_0 b^t$, where b is some positive number and R_0 is the value of $R(t)$ at $t = 0$ (because $R(0) = R_0 \times b^0 = R_0 \times 1 = R_0$). We'll use this form of the exponential equation in the following application also.

ROTTING LEFTOVERS

Any quantity—say a population of rabbits, or of people—that grows by a fixed percentage each year, no matter how small, will double over some period of time. It will then double again after an equal period of time, and so on forever. Every exponentially growing quantity grows in this way, so every exponentially growing quantity is said to have a "doubling time."

Suppose you leave a dish of food out at midnight. The dish happens to have 10 bacteria sitting on it. Suppose also that this population of bacteria increases by 4% every minute. How long will it be before the number of bacteria in the dish doubles? And how many bacteria will you be consuming when you finish off the leftovers at noon the next day?

If the population is growing by 4% every minute, then it is growing exponentially, and can be described by an exponential equation of the form $R(t) = R_0 b^t$, just like the rabbit population in the previous example. We already know R_0, the number of bacteria at time $t = 0$; it's 10. But what is b?

Besides the fact that $R_0 = 10$, we also know that the bacterial population at the end of 1 minute, $R(1)$, is 4% greater than at $t = 0$, because we're told that the population is growing by 4% every minute. This fact can be written down as $R(1) = 10 + (.04 \times 10) = 10.4$.

We also know that $R(1)$ must be given by the exponential equation $R(t) = R_0 b^t$ with 10 plugged in for R_0 and 1 plugged in for t, namely, $R(1) = 10b^1 = 10b$. We can now set this expression for $R(1)$ equal to the number found in the previous paragraph: $10b = 10.04$.

Dividing both sides by 10 to solve for b, we find that $b = 1.004$.

We now have both R_0 and b, and so can write down the exact exponential equation that describes this bacterial population: $R(t) = R_0 b^t = 10 \times (10.004^t)$.

Population Growth

The human race has inhabited Earth for about 3 million years. For much of that time, the world's population was constant at about 10 million people. Life was difficult; most babies died, and people reached old age and usually died by 30. For food, people harvested wild plants and hunted animals.

With the invention of farming and cities about 10,000 years ago, larger local populations became possible. During the first century A.D, some 2,000 years ago, the world's population had grown to about 300 million people. Around the year 1600, as modern science and technology started to come into being, and population began to grow faster. By 1800, there were about 1 billion people on Earth. It took about 3 million years for the world to gain its first billion people, and only 130 years to gain its second billion. Today, there are over 6 billion people living on Earth.

The most common mathematical model for population growth is the exponential function, $Q(t) = {}_0k^t$ (see main text). As the population figure shows the approximate recent exponential growth of world population, the data becomes visible. Dots are actual world population at 1, 2, 3, 4, 5, and 6 billion; the smooth curve is exponential function $2.3236 \times 10^{-10} \times 1.0124^t$, where t is in years A.D.

The world's population will continue to grow for the near future. But it is physically impossible for the Earth's population to continue to grow exponentially, as there is a finite amount of space and potential for growing food.

We can now return to our first question: What is the doubling time? Let us call that unknown number of minutes T_D. Because we start out with 10 bacteria, the number of bacteria after the first doubling time will be 20 (double). So the population at time T_D is given by $R(T_D) = 20 = 10 \times 1.004_{TD}$. To solve for T_D, we must "take the logarithm" of both sides of this equation. Taking the logarithm undoes the exponential operation much the way that subtraction undoes addition or division undoes multiplication (see chapter on Logarithms). We find that $T_D = 173.63$ minutes (about 2 hrs 54 minutes).

The equation $R(t) = 10 \times (10.004^t)$ also tells us how many bacteria you'll be eating at noon the next day. All we need to know is the number of minutes between midnight and noon, which is 12 hours times 60 minutes per hour, or 720 minutes. Thus, $R(720) = 10 \times 1.004^{720}$, or about 177 bacteria.

Not bad, really. As you read this, your intestines contain trillions of bacteria. But remember the power of exponential growth. After 24 hours there will be 3,137 bacteria; after 48 hours, 984,205 bacteria; and after 3 days, 3.088×10^8 bacteria, about as many people as there are in the United States. Look out for a stomach upset—or worse.

EXPONENTS AND EVOLUTION

Predators and sickness often keep populations from growing exponentially, but if they don't there is one thing that certainly will: hunger. There can never be an infinite food supply—even if you could somehow turn the whole Earth into a ball of food, it would be limited. Therefore, sooner or later, any exponentially growing population must run out of food and either stop breeding or start starving.

This principle was first clearly explained English economist and minister Thomas Robert Malthus (1766–1834). In his 1798 book, *An Essay on the Principle of Population*, Malthus wrote: "Population, when unchecked, increases in a geometrical ratio [exponentially]. Subsistence [that is, food supply] increases only in an arithmetical ratio [like a straight line]. A slight acquaintance with numbers will show the immensity of the first power in comparison of the second. . . . This implies a strong and constantly operating check on population from the difficulty of subsistence."

By "a slight acquaintance with numbers" Malthus meant a knowledge of exponents. Population increases exponentially ("in a geometrical ratio") whenever it can, rising in a curve that gets ever steeper; but food supply increases (if it increases at all, say by the clearing and planting if more cropland) approximately as a straight line, that is, according to a "linear" function or "arithmetical ratio." And any exponential function will eventually outrun any linear function. Accordingly, any freely-breeding population must eventually outrun its food supply.

Malthus was talking about the human population, but his logic applies to any biological population. Two English biologists, Charles Darwin (1809–1882) and Alfred Russel Wallace (1823–1913)—realized this when they read Malthus's book in the mid-1800s. Both Darwin and Wallace were trying to think of a mechanism to explain biological evolution, the process whereby new species of animals and plants arise from older ones. People had been suggesting theories of evolution for years, but none of them could explain why evolution happened. However, Malthus's reasoning about population growth

Population Growth

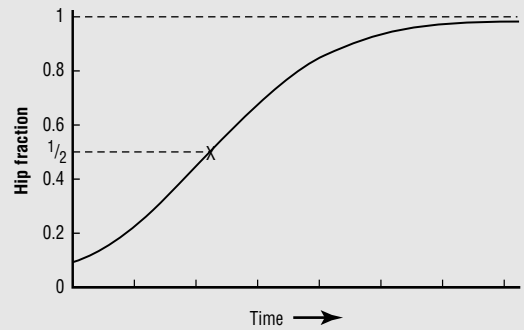
Think of how a rumor spreads. Somebody starts a rumor by telling everyone they know, then those people tell everyone they know, then those people tell everybody they know, and so on.

In a finite environment, such as a high school (or the planet Earth, for that matter), there are a limited number of people to tell. Soon, people who have heard the rumor are having trouble finding people who haven't heard it yet. What happens then?

The function that describes the growing number of people who have heard a rumor is called the logistic curve.

The horizontal axis here is time, the vertical axis the fraction of the school that knows the rumor; let's call it the hip fraction. The curve starts out at time zero at some nonzero number, namely, the fraction of the school population that knows the rumor to begin with. As time goes on, the hip fraction approaches 1; everybody knows the rumor. Using calculus, we can show that the derivative of the logistic curve always has a maximum where the hip fraction equals .5 (marked X on the curve). That is, the rate of change of the hip fraction decreases after that time. A rumor, therefore, spreads more slowly once it has been heard by half the people in a group.

Because germs, like rumors, spread by contact, mathematicians also use the logistic curve to describe the spread of a disease in a finite population.



Logistics curve.

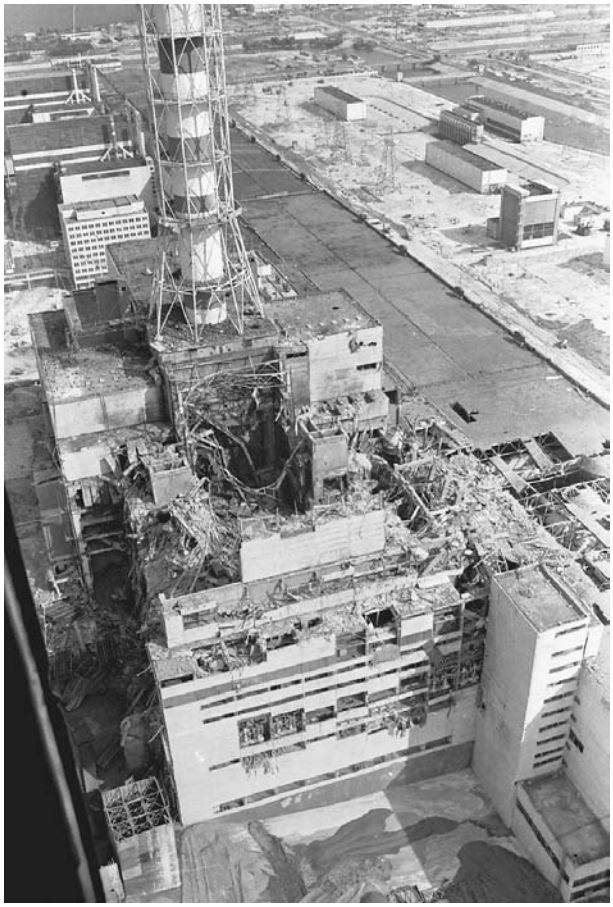
triggered a fresh insight for both Wallace and Darwin. Working separately, they realized that the potential of every species for exponential population growth guaranteed struggle between organisms. In biology, "struggle" usually means not fighting, but competition to leave more offspring than one's rivals. In an article published jointly with Wallace in 1858, Darwin said, echoing Malthus: "[T]he amount of food for each species must, on an average, be constant, whereas the increase of all organisms tends to be geometrical, and in a vast majority of cases at an enormous ratio . . . . Now, can it be doubted, from the struggle each individual has to obtain subsistence, that any minute variation in structure, habits, or instincts, adapting that individual better to the new conditions, would tell upon its vigour and health? In the struggle it would have a better chance of surviving; and those of its offspring which inherited the variation, be it ever so slight, would also have a better chance. . . . Let this work of selection on the one hand, and death on the other, go on for a thousand generations, who will pretend to affirm that it would produce no effect . . . "

Wallace and Darwin's insight was that competition (made inevitable by exponential population growth) is more than just a check or limit on population: it forces nature to filter the chance changes that show up in every generation of creatures and so acts as a creative force, helping sculpt such marvels as the gull's wing, the eagle's eye, and the human brain.

RADIOACTIVE DECAY

In Nevada, about 90 miles (145 km) northwest of Las Vegas, stands an unremarkable-looking ridge of dry, brown rock, owned by the Federal government and known as Yucca Mountain. This is where the United States government hopes to bury 77,000 tons (69,853 tonnes) of highly radioactive nuclear waste from around the U.S. (about 60% of the total amount that had built up as of 2004).

This waste is what is left over when a nuclear power plant has finished using uranium fuel to produce electricity. Plans call for it to be mixed with molten glass and cooled to a solid ("vitrified"), sealed inside rust-resistant metal containers, and parked along 73 miles (117 km) of branching tunnels located 1,000 feet (305 m) below the surface of Yucca Mountain. When the tunnels are full, they will be sealed off and hopefully not entered again—especially by water, which might carry the waste back to the surface—for at least 10,000 years.



A common example of exponential decay is radioactive decay. A concrete sarcophagus covered the damaged nuclear reactor No. 4 at the power plant in Chernobyl, Ukraine, following the 1986 nuclear accident. A/P WIDE WORLD. REPRODUCED BY PERMISSION.

Why does the waste need to be put in such a special place at all? And why for as long as 10,000 years—or for only 10,000 years? Why not forever?

The Reason is Radioactive Decay A "radioactive" substance is one whose individual atoms break apart (fission) at random (chance) times, releasing energy. This energy takes the form of fast-moving particles or invisible rays that can cause cancer or other sickness. Some radioactive atoms are mixed naturally with the environment, whereas some are human-made. Regardless of where they come from, the fewer radioactive atoms we come in contact with, the better for our health. (Some medical tests and treatments do use radioactive substances, however, where the gain is thought to be larger than the risk.)

Radioactive substances disappear over time as their atoms change into other types of atoms. This change in atoms is called "radioactive decay." Like the curve in Figure 2, radioactive decay can be described by an exponential equation with a base between 0 and 1. Someday, therefore, all the nuclear waste generated today will be harmless, but that date is tens or hundreds of thousands of years in the future.

Just as every exponentially increasing process has a doubling time, so every exponentially decaying process has a halving time. In the case of a radioactive substance, this halving time is called the substance's "half-life." Half the atoms in a lump of any radioactive substance will have changed into other substances after one half-life of that substance. Different radioactive substances have different half-lives. Half-lives vary from a tiny fraction of a second up to billions of years.

Consider the substance plutonium 238. Plutonium 238 (also written ^{238}Pu) is both poisonous and radioactive. It can be used as fuel for nuclear reactors or to make nuclear bombs. It is one of the ingredients in radioactive waste of the type that may someday be stored beneath Yucca Mountain (perhaps starting in 2010). ^{238}Pu has a half-life of about 25,000 years. If the amount of ^{238}Pu that we start out with at time $t = 0$ is Q_0 tons, then the amount at some later time t will be described by the exponential equation $Q(t) = Q_0k^t$. Because we know that after the first 25,000 years there will be half as much ^{238}Pu as at time $t = 0$, we can write the following:

$$\frac{Q_0}{2} = Q_0k^{25,000}$$

Because Q_0 appears on both sides of the equal sign, it cancels out. We can then solve for k using logarithms. We find that $k = .999972$. The radioactive decay of 1 ton of ^{238}Pu is shown in Figure 4.

In Figure 4, the radioactive decay of 1 ton of plutonium 238 (^{238}Pu). Time is shown in units of half-lives. Notice that at $t = 1$ (one half-life) the amount of ^{238}Pu is down to

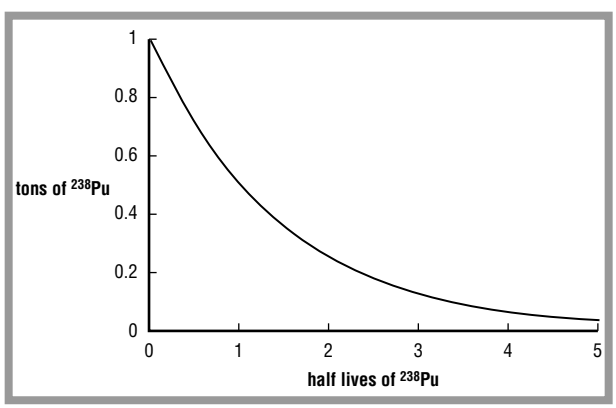


Figure 4.

1/2 ton. To read the time axis in units of years, multiply by 25,000. Note that this curve is exactly the same (except for vertical scale) as the part of Figure 2 to the right of $x = 0$.

As a rule of thumb, experts often say that we should wait at least 10 half-lives of a radioactive substance before considering that a chunk of it is harmless. In the case of plutonium, 10 half-lives are about 250,000 years. How much of an original quantity of any radioactive substance, say 1 ton, is left after 10 half lives? Since waiting one half-life cuts the amount in half, and waiting two half-lives cuts that amount in half, and so on, we can use exponents. A little thought shows that the fraction that is left after ten half-lives is

$$\left(\frac{1}{2}\right)^{10} = .0009766$$

That is, after 10 half-lives, only .0009766 tons (about one one-thousandth of the original ton) will be left.

The wastes intended for Yucca Mountain contain many radioactive substances besides plutonium. Many of these have much shorter half-lives than plutonium, so in about 300 years 99% of the radioactivity of the nuclear waste will have disappeared. After 10,000 years, the amount of time that the Yucca Mountain storage tunnels are supposed to be guaranteed for, these shorter-lived substances will be essentially gone. On the other hand, after 10,000 years only about one-fourth of the plutonium will be gone.

RADIOACTIVE DATING

"Radioactive dating" is an essential technique in geology and archaeology. By looking at the amounts of radioactive substances embedded in rocks or other objects and at the amounts of breakdown products (substances left over from radioactive decay) that are mixed with them, scientists can tell how much radioactive material must have been originally present in the object—and thus, by the exponential equation, how old the object is. For example, if a rock contains 1 gram of radioactive uranium 238 (^{238}U) and 1 gram of lead, which is a breakdown product, it is probable that the rock originally contained 2 grams of ^{238}U. Since ^{238}U has a half-life of about 4.5 billion years, it takes about 4.5 billion years for 2 grams to decay to 1 gram, so we deduce that this particular rock is about 4.5 billion years old.

In practice, breakdown products and radioactive dating are more complex than this. Scientists must look at many different samples of rock (or wood, or whatever material they are dating) and at a number of different radioactive substances and breakdown products, rather than just one. But by combining methods and measuring many different objects, error can be minimized. Through radioactive dating, scientists have verified that the Earth, the Moon, and most meteorites are about 4.5 billion years old. That is, about half the ^{238}U that was present when the Solar System formed has turned into other elements. Some of it, in fact, has ended up in your car battery.

INTEREST AND INFLATION

Let's say that you earn $100 at a weekend job. Your parents insist that you put it in a nice, safe bank until you're 18. They try to comfort you with the idea that your money will earn interest, so you will end up with more money if you wait. How much more? And what exactly does it mean to "earn interest," anyway?

"Interest" is money that is paid to you by a bank in which you have deposited money. The bank invests the money in enterprises that it thinks will be worth more in the future. Banks make profit by taking in more money on their investments than they pay in interest to their depositors (that's you), but regardless of how well a bank's investments are doing, it is obliged to pay you the agreed-upon interest.

To pay interest, the bank looks at the money in your account at regular intervals, say every three months. It then calculates a fixed percentage of that amount (your interest) and adds this money to your account. (At the words "regular intervals" and "fixed percentage" your ears should prick up: "Regular intervals? Fixed percentage? My money will grow exponentially ?" Yes, but wait.) The percentage used to calculate your interest is called the "interest rate."

After another three months (or whatever the interval happens to be), the bank calculates the fixed percentage again and adds it to your account. The interest from the previous time interval—also called a "conversion period"—earns interest during the next time interval, assuming that you haven't taken any money out. This arrangement, where interest earns interest, is called "compound" interest.

Let's go back to your $100. Assume the conversion period is three months (which is one quarter of a year, so it's also called a "quarter"). You get a quarterly interest rate of 1.5%, so the end of the first quarter, the bank adds 1.5% of $100 to your account, namely, $1.50. Your account now contains $101.50. At the end of the second quarter, the bank calculates 1.5% of $101.50, which is $1.52 (rounding down), and it adds that to your account. Your account now contains $103.02. Notice that the amount of interest you receive at the end of the second quarter is larger than the interest you receive at the end of the first quarter. The reason is that you've begun to earn interest on your interest.

Not surprisingly, this is an exponential process. Its equation is $S(n) = P(1 + r)^n$. Here $S(n)$ is the amount of

money in your account after n quarters, P is your principal (the money you start off with, in this case \$100), and r is the quarterly interest rate (1.5%, in this case). Since time, n, is in the exponent, this is an exponential function. Putting in our numbers for P, r, and n, we find that $S(n) = 100\ (1 + .015)^n = 100 \times 1.015^n$.

For the end of the second quarter, $n = 2$, this gives the result already calculated: $S(2) = \$103.02$.

This equation for $S(n)$ should look familiar. It has the same form as the equation for a growing population, $R(t) = R_0 b^t$, with R_0 set equal to \$100 and b set equal to 1.015.

If \$100 is put in the bank when you're 14, then by the time you're 18, four years or 16 quarters later, it will have grown exponentially to $\$100 \times 1.015^{16} = \$126.90$ (rounded up). If you had invested \$1,000, it will have grown to \$1,268.99. That's lovely, but meanwhile there's inflation, which is exponentially making money worth less over time.

Inflation occurs when the value of money goes down, so that a dollar buys less. As long as we all get paid more dollars for our labor (higher wages), we can afford the higher prices, so inflation is not necessarily harmful. Inflation is approximately exponential. For the decade from 1992 to 2003, for example, inflation was usually around 2.5% per year. This is lower than the 6% per year interest rate we've assumed for your invested money, so your \$100 of principal will actually gain buying power against 2.5% annual inflation, but not as quickly as the raw dollar figures seem to show: after four years, you'll have 26% more dollars than you started with (\$126.90 versus \$100), but prices will be 10.4% higher (i.e., something that cost \$100 when you were 14 will cost about \$110 when you are 18).

Furthermore, 6% is a rather high rate for a savings account: during the last decade or so, interest rates for savings accounts have actually tended to be lower than inflation, so that people who keep their money in interest-bearing savings accounts have actually been losing money! This is one reason why many people invest their money in the stock market, where it can keep ahead of inflation. The dark side of this solution is that the stock market is a form of gambling: money invested in stocks can shrink even faster than money in a savings account, or disappear completely. And sometimes it does.

CREDIT CARD MELTDOWN

When you deposit money in a bank, the bank is essentially borrowing your money, and pays you interest for the privilege of doing so. When you borrow money from a bank, you pay the bank interest, so if you don't pay off your debt, it can grow exponentially. Exponential interest growth is why credit-card debt is dangerous. A credit-card interest rate, the percentage rate at which the amount you owe increases per unit time, is much higher than anything a bank will pay to you. (Fifteen percent would be typical, and if you make a late payment you can be slapped with a "penalty rate" as high as 29%.) So if you only make the minimum monthly payments, your debt climbs at an exponential rate that is faster than that of any investment you can make. This is why you can't make a living by borrowing money on a credit card and investing it in stocks. If you could, the economy would soon collapse, because everyone would start doing it, and an economy cannot run on money games; it needs real goods and services.

Those high credit-card interest rates are also the reason credit-card companies are so eager to give credit cards to young people. They count on younger borrowers to get carried away using their cards and end up owing lots of fat interest payments. And it seems like a good bet. In 2004, the average college undergraduate had over \$1,800 in credit-card debt.

The good news is that to avoid high-interest credit-card debt, you need only pay off your credit card in full every month.

THE AMAZING EXPANDING UNIVERSE

The entire Universe is shaped by processes that are described by exponents.

All the stars and galaxies that now speckle our night sky, and all other mass and energy that exists today, were once compressed into a space much smaller than an atom. This super-tiny, super-dense, super-hot object began to expand rapidly, an event that scientists call the Big Bang. The Universe is still growing today, but at different times in its history it has expanded at different speeds. Many physicists believe that for a very short time right after the Big Bang, the size of the Universe grew exponentially, that is, following an equation approximately of the form $R(t) = Ka^t$, where $R(t)$ is the radius of the Universe as a function of time and t and K and a are constants (fixed numbers). This is called the "inflationary Big Bang" theory because the Universe inflated so rapidly during this exponential period. If the inflationary theory is correct, the Universe expanded by a factor of at least 10^{35} in only 10^{-32} seconds, going from much smaller than an electron to about the size of a grapefruit.

This period of exponential growth lasted only a brief time. For most of its 14-billion year history, the Universe's rate of expansion has been more or less proportional to time raised to the 2/3 power, that is, $R(t) = Kt^{2/3}$. Here $R(t)$ is the radius of the universe as a function of time, and K is a fixed number.

Most scientists argue that the Universe will go on expanding forever—and that it's expansion may even be accelerating slowly.

WHY ELEPHANTS DON'T HAVE SKINNY LEGS

The two most common exponents in the real world are 2 and 3. We even have special words to signify their use: raising a number to the power of 2 is called "squaring" it, while raising it to the power of 3 is called "cubing" it. These names reflect the reasons why these numbers are so important. The area of a square that is L meters on a side is given by $A = L^2$, that is, by "squaring" L, while the volume of a cube that is L meters on a side is given by $V = L^3$, that is, by "cubing" L.

These exponents—2 and 3—appear not only in the equations for the areas and volumes of squares and cubes, but for any flat shapes and any solid shapes. For example, the area of a circle with radius L is given by $A = \pi L^2$ and the volume of a sphere with radius L is given by $4/3\ \pi L^3$. The equations for even more complex shapes (say, for the area of the letter "M" or the volume of a Great Dane) would be even more complicated, but would always include these exponents somewhere—2 for area, 3 for volume. We say, therefore, that the area of an object is "proportional to" the square of its size, and that its volume is proportional to the cube of its size.

These facts influence almost everything in the physical world, from the shining of the stars to radio broadcasting to the shapes of animals' legs. The weight of an animal is determined by its volume, since all flesh has about the same density (similar to that of water). If there are two dogs shaped exactly alike, except that one is twice the size of the other, the larger dog is not two times as heavy as the smaller one but 2^3 (eight) times as heavy, because its volume is proportional to the cube of its size. Yet its bones will not be eight times as strong. The strength of a bone depends on its cross-sectional area, that is, the area exposed by a cut right through the bone. The bigger dog's bones will be twice as wide as the small dog's (because the whole dog is twice as big), and area is proportional to the square of size, so the big dog's bones will only be 2^2 (four) times as large in cross section, therefore only four times as strong. To be eight times as strong as the small dog's bones, the big dog's bones would have to be the square root of 8, or about 2.83 times wider.

You can probably see where this is leading. An elephant is much bigger than even a large dog (about ten times taller). Because volume goes by the cube of size, an elephant weights about $10^3 = 10 \times 10 \times 10 = 1000$ times as much as a dog. To have legs that are as strong relative to its weight as a dog's legs are, an elephant has to have leg bones that are the square root of 1,000, or about 31.62 times wider than the dog's. So even though the elephant is only 10 times taller, it needs legs that are almost 32 times thicker. If an elephant's legs were shaped like a dog's, they would snap.

Where to Learn More

Books

Durbin, John R. *College Algebra*. New York: John Wiley & Sons, 1985.

Morrison, Philip, and Phylis Morrison. *Powers of Ten: A Book About the Relative Size of Things in the Universe and the Effect of Adding Another Zero*. San Francisco: Scientific American Library, 1982.

Periodicals

Curtis, Lorenzo. "Concept of the exponential law prior to 1900," *American Journal of Physics* 46(9), Sep. 1978, pp. 896–906 (available at <http://www.physics.utoledo.edu /~ljc /explaw.pdf>.

Wilson, Jim. "Plutonium Peril: Nuclear Waste Storage at Yucca Mountain," *Popular Mechanics*, Jan. 1, 1999.

Web sites

Population Reference Bureau. "Human Population: Fundamentals of Growth: Population Growth and Distribution." <http://www.prb.org/Content/NavigationMenu/PRB/Educators/Human_Population/Population_Growth/Population_Growth.htm> (April 23, 2004).

Factoring

Overview

Factoring a number means representing the number as the product of prime numbers. Prime numbers are those numbers that cannot be divided by any smaller number to produce a whole number. For instance, 2, 3, 5, 7, 11, and 13 (among many others) cannot be divided without producing a remainder.

Factoring in its simplest form is the ability to recognize a common characteristic or trait in a group of individuals or numbers which can be used to make a general statement that applies to the group as a whole.

Another way to think of factoring is that every individual in the group shares something in particular. For example, whether someone is from France, Germany, or Austria is irrelevant in the statement that they are European, because all three of these countries share the geographic characteristic of being on the continent of Europe. The factor that can be applied to all three individuals in this particular group is that they are all European. The ability to recognize relationships between individual components is fundamental to mathematics. Factoring in mathematics is one of the most basic but important lessons to learn in preparation for further studies of math.

Fundamental Mathematical Concepts and Terms

A number which can be divided by smaller numbers is referred to as a composite number.

Composites can be written as the product of smaller primes. For example, 30 has smaller prime numbers which can be multiplied together to achieve the product of 30. These numbers are as follows: $2 \times 3 \times 5 = 30$. A number is considered to be factored when all of its prime factors are recognized. Factors are multiplied together to yield a specific product.

It is important to understand a few basic principals in factoring before further discussion can continue on how factoring can be applied to real life. One of the most important studies of mathematics is to study how individual entities relate to one another.

In multiplying factors which contain two terms, each term must be multiplied with each term of the second set of terms. For example, in (a+b) (a+b), both the a and b in the first set must be multiplied by the a and b in the second set. The easiest way to accomplish this is by employing the FOIL method. FOIL refers to the order of multiplication: first, outer, inner, and last. First we multiply a

by a to yield a^2, then the Outer terms of a and b to yield ab, then the Inner terms of b and a to yield another ab, finally we multiply the Last terms of b and b for b^2. Putting all of these together, we achieve $a^2 + 2ab + b^2$.

Greatest common factor (GCF) refers to two or more integers where the largest integer is a factor of both or all numbers. For example, in 4 and 16, both 2 and 4 are factors that are common to each. However, 4 is greater than 2, so therefore 4 is the greatest common factor. In order to find the greatest common factor, you must first determine whether or not there is a factor that is common to each number. Remember that common factors must divide the two numbers evenly with no remainders. Once a common factor is found, divide both numbers by the common factor and repeat until there are no more common factors. It is then necessary to multiply each common factor together to arrive with the greatest common factor.

Factoring perfect squares is one of the essentials of learning factoring. A perfect square is the square of any whole number. The difference between two perfect squares is the breaking of two perfect squares into their factors. For example $a^2 - b^2$ is referred to as the difference between two perfect squares. The variables a and b refer to any number which is a perfect square. In order to factor $a^2 - b^2$, we must see that the factors must contain both a and b. If we start with $(a - b)$, and remove this expression from $a^2 - b^2$, we will have $(a - b)$ remaining. This would yield a solution of $(a - b)(a - b)$. Using the FOIL method, the product would be $a^2 - ab - ab + b^2$, which is $a^2 - 2ab + b^2$ which is incorrect due to the presence of a middle term.

Alternatively, if we choose $(a + b)$ and remove both a and b from the original equation, we have: $(a + b)(a - b)$. Multiplying these factors back together yields $a^2 - ab + ab - b^2$ which simplifies to our original equation of $(a^2 - b^2)$. The difference between two perfect squares always has alternating + and − signs to eliminate the middle term.

Real-life Applications

Factoring is used to simplify situations in both math and in real life. They allow faster solutions to some problems. In the mathematical calculations used to model problems and derive solutions, factoring plays a key role in solving the mathematics that describe systems and events.

IDENTIFICATION OF PATTERNS AND BEHAVIORS

By learning the patterns and behaviors of factors in mathematical relationships, it is possible to identify similarities between multiple components. By being able to quickly and accurately find similarities, a solution can usually be identified. The solution to any given problem is based on how each individual player or factor in the problem relates to one another for an effective solution. By being able to see these relationships, many times it is possible to see the solution in the relationship.

An example is commonly found in decision making. For example, a shopper enters an unfamiliar grocery store looking for Gouda cheese. The shopper could wander aimlessly, hoping to spot the cheese, but a smarter approach illustrates the intuitive process of factoring. Granted, with enough time, the shopper might eventually find the cheese, but a better approach is to search for a common factor to help narrow the search. What common factor does cheese have with other items in the store? The obvious choice would be to look for the dairy section and eliminate all other sections in the store. The shopper would then further factor the problem to locate the cheese section and eliminate the milk, eggs, etc. Finally one would only look at the cheese selections for the answer, the Gouda cheese. This is a fairly simple nonmathematical example, but it demonstrates the principle of mathematical factoring—a search for similarities among many individual numerical entities.

REDUCING EQUATIONS

In math, one of the most useful applications of factoring is in eliminating needless calculations and terms from complex equations. This is often referred to as "slimming down the equation." If you can find a factor common to every term in the equation, then it can be eliminated from all calculations. This is because the factor will eventually be eliminated through the calculation and simplification process anyway. An example of this is $(2+8)/4$ which can be slimmed down to $(1 + 4)/2$ by eliminating the common factor of 2. The value of the first expression was 10/4 and the value of the second one is 5/2, which is the same once 10/4 is simplified. As we can see, one advantage in eliminating factors is the answer is already simplified. Now let's take a look at a slightly more complicated example:

$$\frac{ax^3 + abx^2 - acx^2}{ax^2}$$

we can see that a common factor of ax^2 can be eliminated.

This expression then becomes:

$$\frac{\cancel{ax^2}(x + b - c)}{\cancel{ax^2}} = (x + b - c)$$

This same technique can be employed in any mathematical equation in which there is a factor common to all parts of the equation.

DISTRIBUTION

Factoring is often used to solve distribution and ordering problems across a range of applications. For example, a simple factoring of 28 yields 4 and 7. In application, 28 units can be subdivided into 4 groups of 7 or 7 groups of 4, Again, by example, in application 28 players could be divided into 4 teams of 7 players or 7 teams of 4 players. This is intuitive factoring—something done every day without realizing that it is a math skill.

SKILL TRANSFER

In addition to factoring mathematical equations, the ability to mathematically factor has been demonstrated to transfer into stronger pattern recognition skills that allow rapid categorization of non-mathematical "factors." Essentially is it an ability to find and eliminate similarities and thus focus on essential difference.

When a defensive linebacker looks over an offensive set in football, he scans for patterns and similarities in numbers of players each side of the ball, in the backfield, in an effort to determine the type of play the opposing quarterback (or his coach) has called. This is not mathematical factoring, but psychology studies have shown that practice in mathematical factoring often leads to a general improvement in pattern recognition and problem solving.

CODES AND CODE BREAKING

Another example of mathematical factoring is in coding and decoding text. Humans have found clever ways of concealing the content of sensitive documents and messages for centuries. Early forms of coding involved the twisting of a piece of cloth over a rod of a certain length. On the cloth would be printed a confusing matrix of seemingly unrelated letters and symbols. When the cloth was twisted over a rod of the proper diameter and length, it would align letters to form messages. The concealed message would be determined by a mathematical factor of proper rod diameter and length that only the intended party would have in possession. Coding and decoding text today is far more complicated. In our new highly computerized age, coding and decoding text depends on an extremely complicated algorithm of mathematical factors.

GEOMETRY AND APPROXIMATION OF SIZE

While factoring is primarily taught and practiced in algebra courses, it is used in every aspect of mathematics. Geometry is no exception. In the field of geometry, there exists the rule of similar triangles. The rule of similar triangles shows that if two triangles have the same angles and the lengths of two legs on one triangle along with a corresponding leg on the other triangle is known, there exists a common factor that can be used to determine the lengths of the other legs. For example, if one wishes to determine the height of a flagpole, factoring through the use of similar triangles can be employed. This is accomplished by an individual of known height standing next to the flagpole. The shadows of both the individual and the flagpole will now be measured. Because the person in standing perpendicular to the ground, a 90-degree triangle is formed with the height of the person being one leg, the length of the shadow being the other leg, and the hypotenuse being the distance from the tip of the person's head to the tip of the head on the shadow. The flagpole forms a similar 90-degree triangle. Once the lengths of the shadows are known, divide the length of the flagpole's shadow by the length of the individual's shadow to determine the common factor. This factor is then multiplied by the height of the individual to find the height of the flagpole.

Potential Applications

In engineering, business, research, and even entertainment, factoring can become a valuable asset. Engineers must use factoring on a daily basis. The job of an engineer is either to design new innovations or to troubleshoot problems as arise in existing systems. Either way, engineers look for effective solutions to complex problems. In order to make their job easier, it is important for them to be able to identify the problem, the solution, and—with regard to the mathematics that describe the systems and events—the factors that systems and events share. Once equations describing systems and events are factored, the most essential elements (the elements that unite and separate systems) can often be more clearly identified. The relationship of each component in the problem will often lead to the solution.

In business, factoring can help identify fundamental factors of cost or expense that impact profits. In research applications, mathematical factoring can reduce complex molecular configurations to more simplified representations that allow researchers to more easily manipulate

Key Terms

Algorithm: A set of mathematical steps used as a group to solve a problem.

Hypotenuse: The longest leg of a right triangle, located opposite the right angle.

Whole number: Any positive number, including zero, with no fraction or decimal.

and design new molecular configurations that result in drugs with greater efficiency—or that can be produced at a lower cost. Factoring even plays a role in entertainment and movie making as complex mathematical patterns related to movement can be factored into simpler forms that allow artists to produce high quality animations in a fraction of the time it would take to actually draw each frame. Factoring of data gained from sensors worn by actors (e.g., sensors on the leg, arms, and head, etc.) provide massive amounts of data. Factoring allows for the simplified and faster manipulation of such data and also allow for mapping to pixels (units of image data) that together form high quality animation or special effects sequences.

Where to Learn More

Web sites

University of North Carolina. "Similar Triangles." <http://www.math.uncc.edu/~droyster/math3181/notes/hyprgeom/node46.html> (February 11, 2005).

AlgebraHelp. "Introduction to Factoring." <http://www.algebrahelp.com/lessons/factoring/> (February 11,2005).

Financial Calculations, Personal

Overview

Unlike calculus, geometry, and many other types of math, basic financial calculations can be performed by almost anyone. These simple financial equations address practical questions such as how to get the most music for the money, where to invest for retirement, and how to avoid bouncing a check. Best of all, the math is real life and simple enough that anyone with a calculator can do it.

Fundamental Mathematical Concepts and Terms

Financial math covers a wide range of topics, broken into three major sections: Spending decisions deals with choices such as how to choose a car, how to load an MP3 player for the least amount of cash, and how to use credit cards without getting taken to the bank; Financial toolbox looks at the basics of using a budget, explains how income taxes work, and walks through the process of balancing a checkbook; Investing introduces the essentials of how to invest successfully, as well as sharing the bottom line on what it takes to retire as a millionaire (almost anyone can do it).

Real-life Applications

BUYING MUSIC

Today's music lover has more choices than ever before. Faced with hundreds of portable players, a dozen file formats, and millions of songs available for instant download, the choices can become a bit overwhelming. These choices do not just impact what people listen to, they can also impact the buyer's finances for years to come. Additionally, in many cases, comparing the different offers can be difficult.

One well-known music service ran commercials during the 2005 Super Bowl, urging music buyers to simply "Do the math" and touting its offer as an unparalleled bargain. The reasoning is that the top-selling music player in 2005 held up to 10,000 songs and allowed users to download songs for about a dollar apiece; buying that player along with 10,000 songs to fill it up would cost around $10,000. But the music service's ad offered a seemingly better deal: unlimited music downloads for just $14.95 per month. While this deal sounds much better, a little math is needed to uncover the real answer.

A good starting point is calculating the "break-even" point: how many monthly payments do we make before we actually spend the same $10,000 charged by the other

firm. This calculation is simple: divide the $10,000 total by the $14.95 monthly fee to find out how many months it takes to spend $10,000. Not surprisingly, it takes quite a few: 668.9 months, to be exact, or about 56 years, which is the break-even point. This result means that if we plan to listen to our downloaded songs for fewer than 56 years, we will spend less with the monthly payment plan. For example, if we plan to use the music for 20 years, we will spend less than $3,600 during that time (20 years × $14.95 per month), a significant savings when compared to $10,000.

One question raised by this ad is, "How many songs does a typical listener really own?" Assuming the user actually does download 10,000 songs, the previous analysis is correct. But 10,000 songs may not be very realistic; in order to listen to all 10,000 songs just one time, a person would have to listen to music eight hours a day for two full months. In fact, most listeners actually listen to playlists much shorter than 10,000 tracks. So if a listener doesn't want all 10,000 tunes, is the $14.95 per month still the better buy?

Again, the calculations are fairly simple. Let's assume we want to listen to music four hours per day, seven days per week, with no repeats each week. By multiplying the hours times the days, we find that we need 28 hours of music. If a typical song is 3 minutes long, then we divide 60 minutes by 3 minutes to find that we need 20 songs per hour, and by multiplying 20 songs by the 28 hours we need to fill, we find that we need 560 songs to fill our musical week without any repeats. Using these new numbers, the break-even calculation lets us ask the original question again: how long, at $14.95 per month, will it take us to break-even compared to the cost of 560 songs purchased outright? In this case, we divide the $560 we spend to buy the music by the $14.95 monthly cost, and we come up with 37.5 months, or just over three years. In other words, at the end of three years, those low monthly payments have actually equaled the cost of buying the songs to start with, and as we move into the fourth and fifth year, the monthly payments begin to cost us more. Plus, for users whose music library includes only 200 or 300 songs, the break-even time becomes even shorter, making the decision even less obvious than before.

Several other important questions also impact the decision, including, "What happens to downloaded music if we miss a monthly payment?" Since subscription services typically require an ongoing membership in order to download and play music, their music files are designed to quit playing if a user quits paying. The result is generally a music player full of unplayable files. A second consideration is the wide array of file formats currently in use. Some services dictate a specific brand of player hardware, while others work with multiple brands. Most users feel that the freedom to use multiple brands offers them better protection for their musical investment. Since some players will play songs stored in multiple formats, they offer users the potential to shop around for the best price at various online stores. A final question deals with musical taste and habits. For listeners whose libraries are small, or who expect their musical tastes to remain fairly constant, buying tracks outright is probably less expensive. For listeners who demand an enormous library full of the latest hits and who enjoy collecting music as a hobby, or for those whose music tastes change frequently, a subscription plan may provide greater value.

Today's music lover has more choices than ever before. Faced with hundreds of portable players, a dozen file formats, and millions of songs available for instant download, the choices can become a bit overwhelming. These choices don't just impact what people listen to, they can also impact the buyer's finances for years to come. KIM KULISH/CORBIS.

In the end, this decision is actually similar to other financial choices involving the question of whether to rent or buy (see sidebar "Rent or Buy?"), since the monthly subscription plan is somewhat like renting music. Math provides the tools to help users make the right choice.

CREDIT CARDS

Although the average American already carries eight credit cards, offers arrive in the mail almost every week encouraging us to apply for and use additional cards. Why are banks so eager to issue additional credit cards to consumers who already have them? Answering this question requires an examination of how credit cards work.

In its simplest possible form, a credit card agreement allows consumers to quickly and easily borrow money for daily purchases. Typically, we swipe our card at the store, sign the charge slip or screen, and leave with our goods. At this point in the process, we have our merchandise, paid for with a "loan" from the credit card issuer. The store has its money, less the fee it paid to the credit card company, and the credit card has paid our bill in exchange for a 2–3% fee and for a promise of payment in full at a later date. At the end of the month, we will receive a statement, pay the entire credit card bill on time to avoid interest or late charges, and this simplest type of transaction will be complete.

If this transaction were the norm, very few companies would enter the credit card business, as the 2–3% transaction fees would not offset their overhead costs. In reality, a minority of consumers actually pay their entire bills at the end of the month, and any unpaid balances begins accruing interest for the credit card issuer. These interest charges are where credit card companies actually earn their profits, as they are, in effect, making loans to thousands of consumers at rates that typically run from 9–14% for the very best customers, from 16–21% for average borrowers, and in the case of customers with poor credit histories, even higher rates. Countless individuals who would never consider financing a car loan or home mortgage at an interest of 16% routinely borrow at this and higher rates by charging various monthly expenses on credit cards, and consequently carrying a balance on their bill.

The average American household with at least one credit card in 2004 carried a credit card balance of $8,400 and as a result paid lenders more than $1,000 in interest and finance charges alone, making the credit card business the most profitable segment of the banking industry today. This fact alone answers the original question of why so many credit cards are issued each year: because they are highly profitable to the lenders. Card issuers mailed out three billion credit card offers in 2004 (an average of ten invitations for every man, woman, and child in the United States) because they know their math: half of all credit card users carry a balance and pay interest, so the more new cards the lenders issue, the greater their profits will be.

Loaning money in exchange for interest is an ancient practice, discussed in numerous historical documents, including the Jewish Torah and the Muslim Koran, which both discuss the practice of usury, or charging exorbitantly high interest rates. Modern U.S. law restricts excessive interest charges, and most states have usury laws on their books that limit the rate that an individual may charge another individual. These rates vary widely from state to state; as of 2005, the usury rate, defined as the highest simple interest rate one individual may legally charge another for a loan, is 9% in the state of Illinois. In contrast, Florida's rate is 18%, Colorado's rate is 45%, and Indiana has no stated usury rate at all. Ironically, these laws do not apply to entities such as pawn brokers, small loan companies, or auto finance companies, explaining why these firms frequently charge rates far in excess of the legal maximums for individuals. Credit card issuers, in particular, have long been allowed to charge interest rates above state limits, making them typically one of the most expensive avenues for consumer borrowing.

How much does it really cost to use credit cards for purchases? The answer depends on several factors, including how much is paid each month and what interest rate is being charged. For this example, we'll assume a credit card purchase of $400, an interest rate of 17%, and a minimum monthly payment of $10. After the purchase and making six months of minimum payments, the buyer has paid $60 (six months $\times$ $10 per month). But because more than half that amount, $33.06 has gone to pay the 17% interest, only $26.94 has been paid on the original $400 purchase. At this point, even though the buyer has paid out $60 of the original bill, in reality $373.06 is still owed ($400 − $26.94).

This pattern will continue until the original purchase is completely paid off, including interest. If the buyer continues making only the required $10 monthly payment, it will take five full years, or 60 payments, to retire the original debt. Over the course of those five years, the buyer will pay a total of $194 in interest, swelling the total purchase price from $400 to almost $600. And if the item originally purchased was an airline ticket, a vacation, or a trendy piece of clothing, the buyer will still be paying for the item long after it's been used up and forgotten. While many factors influence the final cost of saying "charge it," a simple rule of thumb is this: Buyers who pay off their charges over the longest time allowed can expect to pay about 50% more in total cost when putting a purchase on the credit card, pushing a $10 meal to an actual cost of $15. Similarly, a $200 dress will actually cost $300, and a $1,000 trip will actually consume $1,500 in payments.

Credit cards are valuable financial tools for dealing with emergencies, safely carrying money while traveling, and in situations such as renting a car when required to do business. They can also be extremely convenient to use, and in most cases are free of fees for those customers who pay their balance in full each month. Only by doing the math and knowing one's personal spending habits can one know if credit cards are simply a convenient financial tool, or a potential financial time bomb.

CAR PURCHASING AND PAYMENTS

For most consumers, an automobile represents the second largest purchase they will ever make, which makes understanding the car buying process critically important. Several important questions should be considered before buying a new car. First, a potential buyer should calculate how much he can spend. Most experts recommend keeping car payments below 20% of take-home pay, so if a worker receives a check for $2,000 each month (after taxes and other withholding), then he should plan to keep his car payments below $400 (20% × $2,000). This figure is for all car payments, so if he already has a $150 payment for another car, he will be shopping in the $250 per month payment range.

Using this $250 monthly payment, the buyer can consult any of several online payment calculators to determine how much he can spend. For example, if the buyer is willing to spend five years (60 months) paying off his vehicle, this might mean he could afford to borrow about $13,000 for a vehicle (this number varies depending on the actual interest rate at the time of the loan). However this value must pay not just for the car, but also for additional fees such as sales tax, license fees, and registration, which vary from state to state and which can easily add hundreds or thousands of dollars to the price of a new vehicle. For this example, we will estimate sales tax at 6%, license fees at $200, and registration at $100; so a car priced at $12,000 will wind up costing a total of $13,020 (12,000 + .06 × 12,000 + $200 + $100), which is right at the target value of $13,000.

The second aspect of the buying equation is the down payment. A down payment is money paid at the time of sale, and reduces the amount that must be borrowed and financed. In the case of the previous example, a down payment of $2,000 would mean that instead of shopping in the $12,000 price range, the buyer could now shop with $14,000 as the top price.

Many buyers have a used car to sell when they are buying a new vehicle, and in many cases they sell this car to the dealer at the same time, a process known as "trading-in." A trade-in involves the dealer buying a car from the customer, usually at a wholesale price, with the intent to resell it later. A trade-in is a completely separate transaction from the car purchase itself, although dealers often try to bundle the two together. Here again, securing information such as the car's fair trade value will allow the savvy customer to receive a fair price for the trade.

Many consumers find the car-buying experience frustrating, and they worry that they are being taken advantage of. Automobile dealerships are among the only places in the United States where every piece of merchandise has a price tag clearly attached, but both the seller and the buyer know the price on the tag means very little. Most cars today are sold at a significant discount, meaning that a sticker price of $20,000 could easily translate to an actual sales price of $18,000. Incentives, commonly in the form of rebates (money paid back to the buyer by the manufacturer), can chop another $2,000-$5,000 off the actual price, depending on the model and how late in the season one shops. While dealers are willing to negotiate and offer lower prices when they must, they are also going to try to sell at a higher price whenever possible, which places the burden on the buyer to do the homework before shopping. Numerous websites and printed manuals provide actual dealer costs for every vehicle sold in the United States, as well as advice on how much to offer and when to walk away.

CHOOSING A WIRELESS PLAN

Comparing cellular service plans has become an annual ritual for most consumers, as they wrestle with whether to stay with their current cell phone and provider or make the jump to a new company. Beyond the questions of which service offers the best coverage area and which phone is the most futuristic-looking, some basic calculations can help determine the best value for the money.

There are normally three segments to wireless plans. The first segment consists of a set quantity of included minutes that can be used without incurring additional charges. These are typically described as "anytime" minutes, and are the most valuable minutes because they can be used during daytime hours. These minutes are typically offered on a use-it-or-lose-it basis, meaning that if a plan includes 400 minutes and the customer uses only 150, the other 250 minutes are simply lost. Some plans now offer rollover minutes, which means that in the previous example, the 250 minutes would roll to the next month and add to that month's original 400 minutes, providing a total of 650 minutes that could be used without additional charges.

Another segment is that many wireless plans include large blocks of so-called free time, during which calls can be made without using any of the plan's included minutes. These free periods are usually offered during times when the phone network is lightly used, such as late at night and on weekends when most businesses are closed. Users may talk non-stop during these free periods without paying any additional fees.

The third major component of a wireless plan is its treatment of any additional minutes used during non-free periods. In many cases, these additional minutes are

billed at fairly high rates, and using additional minutes past those included in a plan's base contract can potentially double or triple the monthly bill.

Other features are sometimes offered, including perks such as free long-distance calling, premium features such as caller identification, and free voicemail. In other cases, providers allow free calls between their own members as part of so-called affinity plans. Cellular plans are typically sold in one- or two-year contracts.

Choosing a wireless plan can be challenging, since there are so many options, and choosing the wrong plan can be a costly choice. A few guidelines can help simplify this choice. First, users should estimate how many minutes will be needed during non-free periods, and then add 10–15% to this estimate in order to provide a margin of error. Next, users can consider whether an affinity plan or free long distance can impact their choices; in cases where most calls are made between family members, plans with these features can offer significant savings.

Finally, users can compare options among the several providers, paying careful attention to coverage areas. For most users, saving a few dollars per month by choosing a carrier with less coverage winds up being an unsatisfying choice. In addition, users should carefully weigh whether to sign a two-year contract, which may offer lower rates, or a one-year plan. One-year plans provide the most flexibility, since rates generally fall over time and a shorter contract allows one to reevaluate alternative plans more often. In addition, wireless providers are now required to let customers keep their cell numbers when they change providers (a feature called "portability"), simplifying the change-over process.

For users needing very few minutes each month, or those on extremely tight budgets, pay-as-you-go plans offer a thrifty alternative. These plans do not normally include free phones or bundles of minutes; instead, a user recharges the account by buying minutes in credit card form at a convenience store or similar outlet. For users who talk 30 minutes or less each month, these plans can be ideal.

When purchasing a wireless plan, add-ons will inevitably increase the final cost. A plan advertised at $39.95 per month will typically generate bills of $43.00 or more when all the taxes and fees are added in, so plan accordingly.

BUDGETS

Personal budgets fill two needs. First, they measure or report, allowing people to assess how much they are spending and what they are spending on. Second, budgets forecast or predict, allowing people to evaluate where their finances are headed and make changes, if necessary. A budget is much like an annual checkup for finances, and can be simple or complex. The simplest budget consists of two columns, labeled "In" and "Out."

The first step in the budgeting process consists of filling the in column with all sources of income, including wages, bonuses, interest, and miscellaneous income. In the case of income that is received more frequently, such as weekly paychecks, or less frequently, such as a quarterly bonus, one must convert the income to a monthly basis for budget purposes, with quarterly items being divided by three and weekly items being multiplied by four. In the case of semiannual items, such as auto insurance premiums, the amount is divided by six.

Next, in the out column, all identifiable outflows should be listed, such as mortgage/rent payments, utilities (electricity, gas, water), car payments and gasoline, interest expense (i.e., credit card charges), health care, charitable donations, groceries, and eating out. The details of this list will vary from person to person, but an effort should be made to include all expenditures, with particular attention paid to seemingly small purchases, such as soft drinks and snacks, cigarettes, and small items bought with cash. For accuracy, any purchase costing over $1 should be included.

The third step is to add up each column, and find the difference between them; in simplest terms, if the out column is larger than the in column, more money is flowing out than in, the budget is out of balance and the family's financial reserves are being depleted. If more money is flowing in than out, the family's budget is working, and attention should be paid to maintaining this state.

The fourth step in this process is evaluating each of the specific spending categories to determine whether it is consuming a reasonable proportion of the spendable income. For instance, each individual category can be divided by the total to determine the percentage spent; a family spending $700 of their monthly $2,000 on car payments, gas, and insurance should probably conclude that this expenditure (700/2000 = 35%) is excessive and needs to be adjusted. In many cases, families creating a first-time budget find that they are spending far more than they realized at restaurants, and that by cooking more of their own meals they can almost painlessly reduce their monthly deficits.

The previous four steps of this process ask "What is being spent?" The fifth and final step asks, "What should be spent?" or "What is the spending goal?" At a minimum, efforts should be made to bring the entire budget into balance by adjusting specific categories of spending. Ideally,

goals can be set for each category and reevaluated at the end of each month. A budget provides a simple, inexpensive tool to begin taking control of one's personal finances. W. Edwards Deming, the genius who transformed the Japanese from makers of cheap trinkets into the worldwide experts on quality manufacturing, is often paraphrased as saying, "You can't change what you can't measure." A simple three-column budget provides the basic measurement tool to begin measuring one's financial health and changing one's financial future.

UNDERSTANDING INCOME TAXES

The United States Treasury Department collects around $1 trillion in individual income taxes each year from U.S. workers, most of it subtracted from paychecks. While income tax software has taken much of the agony out of tax preparation each April, most workers still have to interact with the Internal Revenue Service, or IRS, from time to time, especially in the area of filling out tax paperwork.

Employers are required by law to withhold money from employee paychecks to pay income taxes. But because each person's tax situation is different, the IRS has a specific form designed to tell employers how much to withhold from each employee. This form, the W-4, asks taxpayers a series of questions, such as how many children they have and whether they expect to file specific tax forms or not. By supplying this form to new employees, companies can ensure that they withhold the proper amount from each paycheck, as well as protect employees from penalties that apply if they do not have enough of their taxes withheld. In cases where family information changes, or where the previous year's withholding amount was too high or too low, a new form can be filed with the employer at any time during the year.

At the end of the calendar year, employers issue a report to each employee called a W-2. Form W-2 is a summary of an employee's earnings for the entire year, including the total amount earned, or gross pay and, amounts withheld for income tax, social security, unemployment insurance, and other deductions. The information from the W-2 is used by the employee when filing federal and state income returns each year. W-2 forms are required to be mailed to employees by January 31; if a W-2 is not received by the first week in February, the employee should contact the employer.

Other forms are used to report other types of income. The 1099 form is similar to W-2s and is sent to individuals who received various types of non-wage income during the year. For example, form 1099-INT is used by banks to provide account holders with a record of interest earned, form 1099-DIV is used to report dividend income, and form 1099-MISC is used to report monetary winnings such as contest prizes, as well as other types of miscellaneous income. These forms should not be discarded, as the amounts on them are reported to the IRS, which matches these reported amounts with individual tax returns to make sure the income was reported and taxes were paid on it. Failure to report income and payroll taxes could lead to penalties and the possibility of a tax audit, in which the taxpayer is required to document all aspects of the tax return to an IRS official.

BALANCING A CHECKBOOK

Balancing a checkbook is an important chore that few people enjoy. A correctly balanced checkbook provides several distinct benefits, including the knowledge of where one's money is being spent, and the avoidance of embarrassing and costly bounced checks. A balanced account also allows one to catch any mistakes, made either by the bank or by the individual, before they create other problems. Balancing a checkbook is actually quite simple and can usually be accomplished in less than half an hour. Whether one uses software or the traditional paper-and-pencil method, the general approach is the same.

Balancing a checkbook begins with good recordkeeping, which means correctly writing down each transaction, including every paper check written, deposit made, ATM withdrawal taken, or check-card purchase made. Bad recordkeeping is a major cause of checkbook balancing problems.

Determining whether all of one's transactions have cleared the checking account is described as the process of a paper check winding its way through the financial system from the merchant to the bank, which can take several days. It also refers to deposits or withdrawals made after the statement date. The net effect of clearing delays is that most consumers will have records of transactions that are not in the latest bank statement, meaning this statement balance may appear either too high or too low. Determining whether all items have cleared involves a review of the records collected in the previous step. A checkmark is placed next to the item on the bank statement for each check, ATM receipt, or other record. Once this process is complete, and assuming good records have been kept, all the items in the bank statement will be checked, and several items that were not in the statement at all will remain. The process of adjusting for these uncleared items is called reconciling the statement.

To reconcile a check register with the bank statement, all the uncleared items must be accounted for, since these transactions appear in the personal check register but not in the statement. Specifically, deposits and other

uncleared additions to the account must be subtracted, while withdrawals, check-card transactions, written checks, and other uncleared subtractions from the account must be added back in. The net effect of this process is to back the records up to the date of the bank statement, at which time the two totals, the check register and the bank statement, should match. Many banks include a simple form on the back of the printed bank statement to simplify this process.

For most customers, a day will arrive when the account simply does not balance. Since bank errors are fairly rare, the most common explanation is an error by the customer. A few simple steps to take include scanning for items entered twice, or not entered at all; data entry errors, such as a withdrawal mistakenly entered as a deposit; simple math errors; and forgetting to subtract monthly service charges or fees. Most balancing errors fall into one of these categories, and as before, good record-keeping will simplify the process of locating the mistake.

Balancing a checkbook is not difficult. The time invested in this simple exercise can often pay for itself in avoided embarrassment and expense.

SOCIAL SECURITY SYSTEM

The Social Security system was established by President Franklin Roosevelt in 1935, creating a national system to provide retirement income to American workers and to insure that they have adequate income to meet basic living expenses. Due largely to this program, nine in ten American senior citizens now live above the official poverty line.

But a Social Security number is important long before one retires. Because the United States does not have an official, government-issued identification program, Social Security numbers are frequently used as personal identification numbers by universities, employers, and banks. U.S. firms are also required by law to verify an applicant's Social Security number as part of the hiring process, making a Social Security card a necessity for anyone wanting to work. For this reason, most Americans apply for and receive a Social Security number and card while they are still minors.

Social Security numbers and cards are issued free of charge at all Social Security Administration offices. An applicant must present documents such as a birth certificate, passport, or school identification card in order to verify the person's identity. After these documents are verified, a number will be issued. A standard Social Security number is composed of three groups of digits, separated by dashes, such as 123-45-6789, and always contains a total of nine digits. Each person's number is unique, and in some cases, the first three digits may indicate the region in which the card was issued. The simplest way for a child to receive a Social Security number is for the parents to apply at birth, at the same time they apply for a birth certificate. After age 12, a child applying for a card, in addition to providing documentation of age and citizenship, must also complete an in-person interview to explain why no card has been previously issued.

When a person begins working, the employer withholds part of the worker's earnings to be deposited into the Social Security system; as of 2005, these contributions are taken out of the first $90,000 in earned income each year at a rate of 7.65%. Starting at age 25, each worker receives an annual statement listing their income for the previous year; this information should be carefully checked for accuracy. While taking one's Social Security card to job interviews or loan applications is a good idea, the Social Security Administration recommends that cards be kept in a safe place, rather than carried on one's person. In the event that a Social Security card is lost or stolen, a new card can be requested at no charge by completing the proper form and submitting verification of identity. The new card will have the same number on it as the old card. In the case of a name change due to marriage, divorce, or similar events, a new card can be issued with the same number and the cardholder's new name. This process requires documentation showing both the previous name and the new name.

The Social Security system remains the largest single retirement plan in the country, is mandatory for most workers, and is expected to remain in place for the foreseeable future.

INVESTING

Investing simply means applying money in such a way that it grows, or increases, over time. In a certain sense, investing is somewhat like renting money to someone else, and in return, receiving a rental fee for the privilege. Investments come in an almost endless variety of forms, including stocks, bonds, real estate, commodities, precious metals, and treasuries. While this array of options may seem bewildering at first, all investment decisions are ultimately governed by a simple principle: "risk equals reward."

Risk is the potential for loss in any investment. The least risky investments are generally government-backed investments, such as Treasury bills and Treasury bonds issued by the United States government. These investments are considered extremely safe because they are backed by the U.S. Treasury and, barring the collapse of the government, will absolutely be repaid. For this reason,

these investments are sometimes described as riskless. At the other end of the risk spectrum might be an investment in a company that is already bankrupt and is trying to pull itself out of insolvency. Because the risk of losing one's investment in such a firm is extremely high, this type of investment is often referred to as a junk bond, since its potential for loss is high. Between riskless and highly risky investments are a variety of other options that provide various levels of risk. Risk is generally considered higher when money is invested for longer periods of time, so short-term investments are inherently less risky than long-term ones.

Reward is the return investors hope to receive in exchange for the use of their money. Most investors are only willing to lend their money to someone for something in return. Investors who buy a rare coin or a piece of real estate are hoping that the value of the coin or house will rise, so they can reap a reward when they sell it. Likewise, investors who buy shares of a company's stock is betting that the company will make money, which it will then pass along to them as a dividend. Investors also hope that as the company grows, other investors will see its value and the stock price itself will rise, allowing them to profit a second time when they sell the stock. Investment rewards take many different forms, but financial returns are the main incentive for people to invest.

The principle "risk equals reward" states that investments with higher levels of risk will normally offer higher returns, while safer (less risky) investments will normally return smaller rewards. For this reason, the very safest investments pay very low rates. An insured deposit in a savings account at a typical U.S. bank earns about 1–2% per year, since these funds are insured and can be withdrawn at any time. Other safe investments, such as U.S. Treasury bills and U.S. savings bonds, pay low interest rates, typically 3–4% for a one-year investment.

Corporate bonds and stocks are two tools that allow public corporations to raise money. Bonds are considered a less risky investment than stocks, and hence pay lower returns, generally a few percentage points higher than Treasury bills. Historically, stocks in U.S. firms have returned an average of 9–10% per year over the long-term. However, this average return conceals considerable volatility, or swings, in value. This volatility means in a given year the stock market might rise by 30-40%, decline by the same amount, or experience little or no change. This variation in annual rates of return is one reason stocks are considered more risky than Treasuries, and hence pay a higher rate of return. Most financial experts recommend that those investing for periods longer than ten years place most of their funds in a variety of different kinds of stocks.

Among the riskiest investments are stock options and commodity futures. Because these types of investments are complex and can potentially lead to the loss of one's entire investment, they are generally appropriate only for experienced, professional investors. Other investments, such as rental real estate, can offer substantial returns in exchange for additional work required to maintain, repair, and manage the property.

A few tricks can help young investors take advantage of certain laws to invest their money. Because the government taxes most forms of income, any investment vehicle that allows the investor to defer (delay) paying taxes will generally produce higher returns with no increase in risk. As an example, consider a worker who begins investing $3,000 per year in a retirement account at age 29. If the worker deposits this money in a normal, taxable savings account or investment fund, each year he will have to pay income tax on the earnings, meaning that his net return will be lower. But if this same amount of money is invested in a tax-sheltered account, the money can grow tax-free, meaning the income each year is higher. Over the course of a career, this difference can become enormous. In this example, the worker's contributions to the taxable account will grow to $450,000 by age 65. But in a tax-sheltered account, those very same contributions would swell to more than $770,000, a 70% advantage gained simply by avoiding tax payments on each year's earnings.

One of the simplest ways to begin a tax-deferred retirement plan is with a Roth Individual Retirement Account (IRA). Available at most banks and investment firms, Roth accounts allow any person with income to open an account and begin saving tax-free. Beginning in 2005, the maximum annual contribution to a Roth IRA is $4,000, which will increase again in 2008 to $5,000. One notable feature of IRAs is the hefty 10% penalty paid on withdrawals made before retirement. While this may seem like a disadvantage, this penalty provides strong incentive to keep retirement funds invested, rather than withdrawing them for current needs.

Another outstanding investment option is a 401(k) plan, offered by many large employers under a variety of names. These plans not only allow earnings to grow tax-deferred like an IRA, they offer other advantages as well. For instance, most firms will automatically withdraw 401(k) contributions from an employee's paycheck, meaning he doesn't have to make the decision each month whether to invest or not. Also, some companies offer to match employee contributions with additional contributions. In a case where a company offers a 1:1 match on the first $2,000 an employee saves, the employee's $2,000 immediately becomes $4,000, equal to

a 100% return on the investment the first year, with no added risk. In the case of a 50% match on the first $3,000, the firm would contribute $1,500. Company matches are among the best deals available and should always be taken advantage of.

Investing is a complex subject, and investing in an unfamiliar area is a chance for losses. By choosing a variety of investments, most investors can generate good returns without exposing themselves to excessive risk. And by taking time to learn more about investment options, most investors can increase their returns without unduly increasing their risk.

RETIRING COMFORTABLY BY INVESTING WISELY

Who wants to be a millionaire? More importantly, what chance does an average 18-year-old person have of actually reaching that lofty plateau? Surprisingly, almost anyone who sets that as a goal and makes a few smart choices and exercises self-discipline along the way can fully expect to be a millionaire by the time he retires. In fact, there are so many millionaires in the United States today that most people already know one or two, even though they are tough to pick out since few of them fit the common stereotype (see sidebar: Millionaire Myths).

Is a million dollars enough to retire comfortably on? Most people would scoff at the question, but the answer may not be as obvious as it first seems. Most members of the World War II generation clearly remember an era of $5,000 houses, $500 cars, and 5-cent soft drinks. What they may not recall so clearly is that in 1951, the average American worker earned only $56.00 per week, meaning that while prices are much higher today, wages have risen substantially as well.

This gradual rise in prices (and the corresponding fall in the purchasing power of a dollar) is called inflation. When inflation is low, and prices and wages increase 3–4% per year, most economists feel the economy is growing at a healthy pace. When inflation reaches higher levels, such as the double-digit rates experienced in the late 1970s, the national economy begins to collapse. And in rare situations, a disastrous phenomena known as hyperinflation takes over. In 1922, Germany experienced an inflation rate of 5,000%. This staggering rate meant that in a two-year period, a fortune of 20 billion German marks would have been reduced in value to the equivalent of one mark. One anecdotal account of hyperinflation in Germany tells of individuals buying a bottle of wine in the expectation that the following day the empty bottle could be sold for more than the full bottle originally cost. Hyperinflation has occurred more recently as well: Peru, Brazil, and Ukraine all experienced hyperinflation during the 1990s; with prices rising quickly, sometimes several times each day, workers began demanding payment daily so they could rush out and spend their earnings before the money lost much of its value.

While hyperinflation can destroy a nation's economy, it is a rare event. A far more realistic concern for workers intent on retiring comfortably is the slow but steady erosion of their money's value by inflation. In the same way that the 5-cent sodas of the 1950s now cost more than a dollar, an increase of twenty-fold, one must assume that the one-dollar sodas of today may well cost $20 by the middle of the twenty-first century. And as costs continue to climb, the value of a dollar, or a million dollars, will correspondingly fall.

The million dollar question (will a million dollars be enough?) can be answered fairly simply using a mathematical approach and several steps. The first question: how much money will be needed in 50 years to equal the value of $1 million today? The first step of this process is determining how much buying power $1 million loses in one year. If the rate of inflation is 3%, a reasonable guess, then over the course of one year $1 million is reduced in buying power by 3%. At the end of the first year, it has buying power equal to $1,000,000 × 97%, or $970,000. This is still a fantastic sum of money to most people, but the true impact of inflation is not felt in the first year, but in the last.

These calculations could continue indefinitely, multiplying $970,000 × 97% to get the value at the end of the second year, and so forth. If this were done for 50 years, we could eventually produce an inflation "multiplier," a single value by which we multiply our starting value to find the predicted future buying power of that sum. In this example, the inflation multiplier is .22, which we multiply by our starting sum of $1 million to find that at retirement in 50 years the nest-egg will have the buying power of only $220,000 today. And while $220,000 is a nice sum of money, it may not be enough to support a comfortable retirement for very many years.

This raises another obvious question: how much will it take in 50 years to retain the buying power of $1 million today? This calculation is basically the inverse of the previous one. To determine how much is required one year hence to have the buying power of $1 million today, we simply multiply by 1.03 (based on our 3% inflation assumption), giving a need next year for $1,030,000. Again, we can carry this out for 50 years and produce a multiplier value, which in this case turns out to be 4.5. We then multiply that value times the base of $1 million to learn that in order to have the buying power of $1 million

today will require one to have accumulated more than $4 million by retirement.

In summary, the answer to the question is simple: If a retirement fund of $220,000 would be adequate for today, then $1 million will be adequate in 50 years. But if it would take $1 million to meet one's retirement needs today, the goal will need to be quite a bit higher, since today's college students will likely retire in an era when a bottle of drinking water will set them back $20.

This example requires that we picture our bank account as a swimming pool and the money we save as water. The goal is to fill the pool completely by the time of retirement. Because the pool begins completely empty, the task may seem daunting. But like most challenging goals, this one can be achieved with the right approach.

In order to fill the pool, one must attach a pipe that allow water to flow in, and the first decision relates to the size of this pipe, since the larger the pipe, the more water it can carry and the faster the pool will fill. The size of the pipe equates to income level, or for this illustration, the total amount we expect to earn over an entire career. This first decision may be the single most important choice one makes on the road to millionaire status, since this first decision will largely determine the size of the pipe and the size of one's income.

Educational level and income are highly correlated, and not surprisingly, less education generally equates to less income. A report by the U.S. Census Office provides the details to support this claim, finding that students who leave high school before completion can expect to earn about $1 million over their careers. While this sounds like a hefty amount, it is far below what most families need to live, and almost certainly not enough to amass a million dollars in retirement savings. Just for comparison, this value equates to annual earnings of less than $24,000 per year. In our current illustration, this equates to a tiny pipe, and means the swimming pool will probably wind up empty.

The good news from the report is that each step along the educational path makes the pipe a little larger, and fills the pool a little faster. For high school students who stay enrolled until graduation, lifetime earnings climb by 20%, to $1.2 million, meaning that a high school junior who chooses to finish school rather than dropping out will earn almost a quarter of a million dollars for his or her efforts. And with each diploma comes additional earning power. An associate's degree raises average lifetime earnings to $1.5 million, while a bachelor's degree pushes average lifetime earnings to $2.1 million, more than double the amount earned by the high school dropout. Master's degrees, doctorates, and professional degrees such as law and medical degrees each raise expected earnings as well, increasing the size of the pipe and filling the pool faster. Simple logic dictates that when the pipe is two to four times as large, the pool will fill far more quickly. For this reason, one of the best ways to predict an individual's retirement income level is simply to ask, "How long did you stay in school?"

Retirement savings are impacted by income level in multiple ways. First, since every household has to pay for basics such as food, housing, clothing, and transportation, total income level determines how much is left over after these expenses are paid each month, and therefore how much is available to be invested for retirement. Second, as detailed in the Social Security system section, Social Security pays retirement wages based on one's earnings while working, so those who earn more during their career will also receive larger Social Security payments after retirement. Third, employers frequently contribute to retirement plans for their workers, and the level of these contributions is also tied directly to how much the worker earns, with higher earnings equating to higher contributions and greater retirement income. Because each of these pieces of the retirement puzzle is tied to income level, each one adds to the size of the pipe, and helps fill the pool more quickly. Again, education is a primary predictor of income level.

Of course a few people do manage to strike it rich in Las Vegas or win the state lottery, which is roughly equivalent to backing a tanker truck full of water up to the pool and dumping it in. For these few people, the size of the income pipe turns out to be fairly unimportant, since they have beaten some of the longest odds around. To get some idea just how unlikely one is to actually win a lottery, consider other possibilities. For example, most people don't worry about being struck by lightning, and this is reasonable, since a person's odds of being struck by lightning in an entire lifetime are about one in 3,000, meaning that on average if he lived 3,000 lifetimes, he would probably be struck only once. And even though shark attacks make the news virtually every year, the odds of being attacked by a shark are even lower, around one in 12,000.

Since most people fully expect to live their entire lives without being attacked by a shark or being struck by lightning, it seems far-fetched that many would play the lottery each week, given that the odds of winning are astronomically worse. As an example, the Irish Lotto game, which offers some of the best odds of any national lottery on the planet, gives buyers a 1-in-5 million chance of winning, meaning a player is 1,600 times more likely to be struck by lightning than to win the jackpot. And the U.S. PowerBall game offers larger jackpots, but even lower

Millionaire Myths

Say the word "millionaire," and most Americans picture Donald Trump, fully decked out in expensive designer suits and heavy gold jewelry. To most Americans, yachts, mansions, lavish vacations, and fine wines are the sure signs that a person has made it big and has accumulated a seven-figure net worth. But recent research paints a very different picture: most millionaires live fairly frugal lives and tend to prefer saving over spending, even after they've made it big. In fact, the most surprising fact about real millionaires is this: they don't look or act at all like TV millionaires.

The average millionaire in the United States today buys clothes at J.C. Penney's, drives an American made car (or a pickup), and has never spent more than $250 on a wristwatch. He or she inherited little or nothing from parents and has built the fortune in such industries as rice farming, welding contracting, or carpet cleaning. This person is frugal, remains married to the first spouse, has been to college (but frequently was not an A student), and lives in a modest house bought 20 years ago.

In short, while most millionaires are gifted with vision and foresight, there is little they have done that cannot be duplicated by any hard-working, dedicated young person today. The basic principles of accumulating wealth are not hard to understand, but they require hard work and self-discipline to apply.

odds of winning: a player in this game is 16 times less likely to win than in the Irish Lotto, meaning the average PowerBall player should expect to be struck by lightning 26,000 times as often as he wins the jackpot. Of all the unlikely events that might occur, winning the lottery is among the most unlikely.

Once the pipe is turned on, which means we have begun making money, one may find the pool filling too slowly, which means assets and savings are accumulating too slowly. At this point it becomes necessary to notice that the pool includes numerous drains in the floor, some large and others small. Water is continually flowing out these drains, which represent financial obligations such as utility bills, tuition payments, mortgages, and grocery costs. In some cases, the water may flow out faster than the pipe can pump it in, causing the water level to drop until the pool runs dry, meaning the employee runs out of money, and bankruptcy follows. In most families, the inflow and outflow of money roughly balance each other, and each month's bills are paid with a few dollars left, but the pool never really fills up. In either case, retirement will arrive with little or nothing saved, and retirement survival will depend largely on the generosity of the Social Security system.

A more pleasant alternative involves closing some of the drains in the pool, or reducing some expenditures. For most families, the largest drains in the pool will be monthly items such as mortgage and car loan payments that are set for periods of several years and may not be easily changed over the short-run. For these items, decisions can only be made periodically, such as when a new car or home is purchased.

However, some seemingly small items may create huge drains in the family financial pool. For most families, eating out consumes a majority of the food budget, even though eating at home is typically both cheaper. Numerous small bills such as cable, wireless, and internet access can add up to take quite a drink out of the pool, even though each one by itself seems small. Yet, while the total dollar value of such items may seem insignificant, their impact over time can be enormous. By removing just $50 from consumption and investing it at 8% each month during the 50 years of a career, this trivial amount will grow to almost $350,000. These types of choices are among the most difficult to make, but can be among the most significant, especially considering that $50 per month represents what many Americans spend on soft drinks or gourmet coffee. A good rule of thumb for this calculation is to multiply the monthly contribution times 7,000 to find its future value at retirement, assuming one begins at age 20 and retires at age 70.

The other major factor in retiring comfortably is time. To put it simply, the final value of one dollar invested at age 20 will be greater than the final value of four dollars invested at age 50. This means that $10,000 invested at age 20 will grow to $143,000 by age 75, while $40,000 invested at age 50 will be worth only $134,000 at the same time. In fact, a good general rule of thumb is for each eight years that pass, the final value of the retirement nest egg will be reduced by 50%. It is never too early to start saving for retirement.

CALCULATING A TIP

After the meal is over and everyone is stuffed, it's time to pay the bill and make one of the most common financial calculations: deciding how much to tip a server. Some diners believe that the term "tips" is an acronym for

"to insure prompt service," hence they believe that the size of the tip should be tied to the level of service, with excellent service receiving a larger tip and poor service receiving less, or none. Others recognize that servers often make sub-minimum wage salaries (as of 2005, this could be as little as $2.13 per hour) and depend on tips for most of their income, hence they generally tip well regardless of the level of service. Another important consideration is that servers are often the victims of kitchen mistakes and delays, and therefore penalizing them for these problems seems unreasonable. A good general rule of thumb is to tip 20% for outstanding service, 15% for good service, and 10% or less for poor service. Regardless of which tipping philosophy one adopts, some basic math will help calculate the proper amount to leave.

For example, imagine that the bill for dinner is $56.97, which includes sales tax. By looking at the itemized bill, we determine that the pre-tax total is $52.75, since most people calculate the tip on the food and drink total, not including tax. Since the service was excellent we choose to tip 20%. Most tip calculations begin by figuring the simplest calculation, 10%, since this figure can be determined using no real math at all. Ten percent of any number can be found simply by moving the decimal point one place to the left. In the case of our bill of $52.75, we simply shift the decimal and wind up with 10% being $5.275, or five dollars twenty seven and one-half cents. Then to get to 20%, we simply double this figure and wind up with a tip of $10.55.

In real life, we are not concerned about making our tip come out to an exact percentage, so we generally round up or down in order to simplify the calculations. In this case, we would round the $5.275 to $5.25, which is then easily doubled to $10.50 for our 20% tip. Finding the amount of a 15% tip can be accomplished either of two ways. First, we can take the original 10% value and add half again to it. In this case, half of the original $5.27 is about $2.50, telling us that our final 15% tip is going to be around $7.75, which we might leave as-is or round up to $8.00 just to be generous. A second, less-obvious approach involves our two previous calculations of 10% and 20%. Since 15% is midway between these two values, we could take these two numbers and choose the midway point (a process that mathematicians call "interpolation"). In other words, 10% is $5.27 (or about $5.00) and 20% is $10.55 (or about $11.00), so the midway point would be somewhere in the $7.00–8.00 range. Either of these two methods will allow us to quickly find an approximate amount for a 15% tip.

A potential customer looks at exchange rates outside an exchange shop in Rome. AP/WIDE WORLD PHOTOS. REPRODUCED BY PERMISSION.

CURRENCY EXCHANGE

Because most nations issue their own currency, traveling outside the United States often requires one to exchange U.S. dollars for the destination nation's currency. But this process is complicated by the fact that one unit of a foreign currency is not worth exactly one U.S dollar, meaning that one U.S. dollar may buy more or fewer units of the local currency. Currency can be exchanged at many banks and at most major airports, normally for a small fee. Banks generally offer better exchange rates than local merchants, so travelers who plan to stay for some time typically exchange larger amounts of money at a bank when they first arrive, rather than smaller amounts at various shops or hotels during their stay.

Consider a person who wishes to travel from the United States to Mexico and Canada. Before leaving the States, the traveler decides to convert $100 into Mexican currency and $100 into Canadian currency. At the currency exchange kiosk, there is a large board that displays various currencies and their exchange rates.

The official unit of currency in Mexico is the peso, and the listed exchange rate is 11.4, meaning that each

Currencies of the European Community. OWEN FRANKEN/CORBIS.

U.S. dollar is worth 11.4 pesos. Multiplying 100 × 11.4, the person learns that one is able to purchase 1,140 pesos with $100. Canadians also use dollars, but Canadian dollars have generally been worth less than U.S. dollars. On the day of the exchange, the rate is 1.3, meaning that each U.S. dollar will buy $1.3 Canadian dollars, so with $100 the person is able to purchase 130 Canadian dollars. At this point, the shopper might wonder about the exchange rate between Canadian dollars and pesos. Since it is known that 130 Canadian dollars equals the value of 1,140 pesos, the person can simply divide 1,140 by 130 to determine that on this date, the exchange rate is 8.77 pesos to one Canadian dollar.

Exchange rates fluctuate over time. On a business trip one year later, this same person might find that the $100 would now buy 2,000 pesos, meaning that the U.S. dollar has become stronger, or more valuable, when compared to the peso. Conversely, it might be that the dollar has weakened, and will now purchase only 800 pesos. These fluctuations in exchange rates can impact travelers, as the changing rates may make an overseas vacation more or less expensive, but they can be particularly troublesome for large corporations that conduct business across the globe. In their situation, products made in one country are often exported for sale in another, and changing exchange rates may cause profits to rise or fall as the amount of local currency earned goes up or down.

In addition to U.S. dollars, other well-known national currencies (along with their exchange rates in early 2005) include the British pound (.52), the Japanese yen (105), the Chinese yuan (8.3), and the Russian ruble (27.7). Beginning in 2002, 12 European nations, including Germany, Spain, France, and Italy, merged their separate currencies to form a common European currency, the Euro (.76). Designed to simplify commerce and expand trade across the European continent, conversion to the Euro was the largest monetary changeover in world history.

Where to Learn More

Books

Stanley, Thomas, and William Danko. *The Millionaire Next Door.* Atlanta: Longstreet Press, 1996.

Key Terms

Balance: An amount left over, such as the portion of a credit card bill that remains unpaid and is carried over until the following billing period.

Bankruptcy: A legal declaration that one's debts are larger than one's assets; in common language, when one is unable to pay his bills and seeks relief from the legal system.

Bouncing a check: The result of writing a check without adequate funds in the checking account, in which the bank declines to pay the check. Fees and penalties are normally imposed on the check writer.

Inflation: A steady rise in prices, leading to reduced buying power for a given amount of currency.

Interest: Money paid for a loan, or for the privilege of using another's money.

Lottery: A contest in which entries are sold and a winner is randomly selected from the entries to receive a prize.

Mortgage: A loan made for the purpose of purchasing a house or other real property.

Reconcile: To make two accounts match; specifically, the process of making one's personal records match the latest records issued by a bank or financial institution.

Register: A record of spending, such as a check register, which is used to track checks written for later reconciliation.

Web sites

A Moment of Science Library. "Yael and Don Discuss Interpreting the Odds." <http://www.wfiu.indiana.edu/amos/library/scripts/odds.html> (March 6, 2005).

Balanced Scorecard Institute. "What is the Balanced Scorecard?" <http://www.balancedscorecard.org/basics/bsc1.html> (March 6, 2005).

Car-Accidents.com. "2004 Statistics." <http://www.car-accidents.com/pages/stats/2000_killed.html> (March 4, 2005).

CNN.com Science and Space. "Scared of Sharks?" <http://www.cnn.com/2003/TECH/science/12/01/coolsc.sharks.kellan/> (March 3, 2005).

Edmunds New Car Pricing, Used Car Buying, Auto Reviews. "New Car Buying Advice." <http://www.edmunds.com/advice/buying/articles/43091/article.html> (March 5, 2005).

Ends of the Earth Training Group. "W. Edwards Deming's Fourteen Points and Seven Deadly Diseases of Management." <http://www.endsoftheearth.com/Deming14Pts.htm> (March 7, 2005).

Euro Banknotes and Coins. "History of the Euro." <http://www.euro.ecb.int/en/what/history.html> (March 5, 2005).

Fidelity Personal Investments. "U.S. Treasury Securities." <http://personal.fidelity.com/products/fixedincome/potreasuries.shtml> (March 4, 2005).

First National Bank of St. Louis. "Roth IRA Calculator." <http://www.fnbstl.com/tools/calcs/tools.asp?Tool=RothIRA> (March 4, 2005).

Internal Revenue Service. "Form W-4 (2005)." <http://www.irs.gov/pub/irs-pdf/fw4.pdf> (March 5, 2005).

Internal Revenue Service. "Treasury Department Gross Tax Collections: Amount Collected by Quarter and Fiscal Year." <http://www.nlm.nih.gov/hmd/about/collectionhistory.html> (March 4, 2005).

Lectric Law Library. "State Interest Rates and Usury Limits." <http://www.lectlaw.com/files/ban02.htm> (March 7, 2005).

The Lottery Site. "Lottery Odds and Your Real Chance of Winning." <http://www.thelotterysite.com/lottery_odds.htm> (March 5, 2005).

Money Savvy: Yakima Valley Credit Union. "Keep Your Checkbook Up to Date." <http://hffo.cuna.org/story.html?doc_id=218&sub_id=tpempty> (March 7, 2005).

Snopes.com Tip Sheet. "Tip is an acronym for To Insure Promptness." <http://www.snopes.com/language/acronyms/tip.htm> (March 5, 2005).

Social Security Online. "Your Social Security Number and Card." <http://www.ssa.gov/pubs/10002.html> (March 5, 2005).

U.S.Info.State.Gov. "A Brief History of Social Security." <http://www.nlm.nih.gov/hmd/about/collectionhistory.html> (March 4, 2005).

William King Server, Drexel University. "Hyperinflation." <http://william-king.www.drexel.edu/top/prin/txt/probs/infl7.html> (March 5, 2005).

XE.com. "XE.com Quick Cross Rates." <http://www.nlm.nih.gov/hmd/about/collectionhistory.html> (March 7, 2005).

Fractals

Overview

A fractal is a kind of mathematical equation of which pictures are frequently made. A small unit of structural information structure forms the basis for the overall structure. The repeats do not have to be exact, but they are close to the original. For example, the leaves on a maple tree are not exactly alike, but they are similar.

The beauty principle in mathematics states that if a principle is elegant (arrives at the answer as quickly and directly as possible), then the probability is high that it is both true and useful. Fractal mathematics fulfills the beauty principle. Both in the natural world and in commerce, fractals are ever-present and useful.

A fractal has infinite detail. This means that the more one zooms in on a fractal the more detail will be revealed. An analogy to this is the coastline of a state like Maine. When viewed from a satellite, the ocean coastline of the state shows large bays and peninsulas. Nearer to the ground, such as at 40,000 feet (12,192 m) in a jet aircraft flying over the state, the convoluted nature of the coast looks similar, only the features are smaller. If the plane is much lower, then the convolutions become even smaller, with smaller bays and inlets visible but still have basically the same shape. A fractal is similar to the example of the Maine coastline. As the view becomes more and more magnified, the never-ending complexity of the fractal is revealed.

Fundamental Mathematical Concepts and Terms

Among the many features of fractals are their non-integer dimensions. Integer dimensions are the whole number dimensions that most people are familiar with. Examples include the two dimensions (width and length; this is also commonly referred to as 2-D) of a square and the three dimensions (width, length and height; commonly called 3-D) of a cube. It is odd to think that dimensions can be in between 2-D and 3-D, or even bigger than 3-D. But such is the world of fractals.

Dimensions of 1.8 or 4.12 are possible in the fractal world. Although the mathematics of fractals involves complex algorithms, the simplest way to consider fractal dimensions is to know that dimensions are based on the number of copies of a shape that can fit into the original shape. For example, if the lines of a cube are doubled in length, then it turns out that eight of the original-sized cubes can fit into the new and larger cube. Taking the log of 8 (the number of cubes) divided by the log of 2 (doubling in size) produces the number three. A cube, therefore, has three dimensions.

For fractals, where a pattern is repeated over and over again, the math gets more complicated, but is based on the same principle. When the numbers are crunched, the resulting number of dimensions can be amazing. For example, a well-known fractal is called Koch's curve. It is essentially a star in which each original point then has other stars introduced, with the points of the new stars becoming the site of another star, and on and on. Doing the calculation on a Koch's curve that results from just the addition of one set of new stars to the six points of the original star produces a dimension result of 1.2618595071429!

BUILDING FRACTALS

Fractals are geometric figures. They begin with a simple pattern, which repeats again and again according to the construction rules that are in effect (the mathematical equation supplies the rules).

A simple example of the construction of a fractal begins with a + shape. The next step is to add four + shapes to each of the end lines. Each new + is only half as big as the original +. In the next step, the + shapes that are reduced by half in size are added to each of the three end lines that were formed after the first step. When drawn on a piece of paper, it is readily apparent that the forming fractal, which consists of ever smaller + shapes, is the shape of a diamond. Even with this simple start, the fractal becomes complex in only a handful of steps. And this is a very simple fractal!

SIMILARITY

An underlying principle of many fractals is known as similarity. Put another way, the pattern of a fractal is the repetition of the same shaped bit. The following cartoons will help illustrate self-similarity.

In Figure 1, the two circles are alike in shape, but they do not conform to this concept of similarity. This is because multiple copies of the smaller circle cannot fit inside the larger circle.

In Figure 2, the two figures are definitely not similar, because they have different shapes.

The two triangles in Figure 3 are similar. This is because four of the smaller triangles can be stacked together to produce the larger triangle. This allows the smaller bits to be assembled to form a larger object.

Figure 1.

Figure 2.

Figure 3.

A Brief History of Discovery and Development

Fractals are recognized as a way of modeling the behavior of complex natural systems like weather and animal population behavior. Such systems are described as being chaotic. The chaos theory is a way of trying to explain how the behavior of very complex phenomena can be predicted, based on patterns that occur in the midst of the complexity.

Looking at a fractals, one can get the sense of how fractals and chaos have grown up together. A fractal can look mind-bendingly complex on first glance. A closer inspection, however, will reveal order in the chaos; the repeated pattern of some bit of information or of an object. Thus, not surprisingly, the history of fractals is tied together with the search for order in the world and the universe.

In the nineteenth century, the French physicist Jules Henri Poincaré (1854–1912) proposed that even a miniscule change in a complex system that consisted of many relationships (such as an ecosystem like the Florida Everglades or the global climate) could produce a result to the system that is catastrophic. His idea came to be known as the "Butterfly effect" after a famous prediction concerning the theory that the fluttering of a butterfly's wings in China could produce a hurricane that would ravage Caribbean countries and the southern United States. The Butterfly effect relied on the existence of order in the midst of seemingly chaotic behavior.

In the same century, the Belgian mathematician P. F. Verhulst (1804–1849) devised a model that attempted to explain the increase in numbers of a population of creatures. The work had its beginning in the study of rabbit populations, which can explosively increase to a point where the space and food available cannot support their numbers. It turns out that the population increase occurs predictably to a certain point, at which time the growth in numbers becomes chaotic. Although he did not realize it at the time, Verhulst's attempt to understand this behavior touched on fractals.

Leaping ahead over 100 years, in 1963 a meteorologist from the Massachusetts Institute of Technology named Edward Lorenz made a discovery that Verhulst's model was also useful to describe the movement of complicated patterns of atmospheric gas and of fluids. This discovery spurred modern research and progress in the fractal field.

In the late 1970s, a scientist working at International Business Machines (IBM) named Benoit Mandelbrot was working on mathematical equations concerning certain properties of numbers. Mandelbrot printed out pictures of the solutions and observed that there were small marks scattered around the border of the large central object in the image. At first, he assumed that the marks were created by the unclean roller and ribbon of the now-primitive inkjet type printer. Upon a closer look, Mandelbrot discovered that the marks were actually miniatures of the central object, and that they were arranged in a definite order. Mandelbrot had visualized a fractal.

This initial accidental discovery led Mandelbrot to examine other mathematical equations, where he discovered a host of other fractals. Mandelbrot published a landmark book, *The Fractal Geometry of Nature,* which has been the jump-start for numerous fractal research in the passing years.

Real-life Applications

FRACTALS AND NATURE

Fractals are more than the foundation of interesting looking screensavers and posters. Fractals are part of our world. Taking a walk through a forest is to be surrounded by fractals. The smallest twigs that make up a tree look like miniature forms of the branches, which themselves are similar to the whole tree. So, a tree is a repeat of a similar (but not exact) pattern. The leaves on a softwood tree like a Douglass fir or the needles on a hardwood tree like a maple are almost endless repeats of the same pattern as well. So are the stalks of wheat that sway in the breeze in a farmer's field, as are the whitecaps on the ocean and the grains of sand on the beach. There are endless fractal patterns in the natural world.

In the art world, the popularity of the late painter Jackson Pollock's seemingly random splashes of color on his often immensely-sized canvasses relate to the fractal nature of the pattern. Pollock's paintings reflected the fractal world of nature, and so strike a deep chord in many people.

By studying fractals and how their step-by-step increase in complexity, scientists and others can use fractals to model (predict) many things. As we have seen above, the development of trees is one use of fractal modeling. The growth of other plants can be modeled as well. Other systems that are examples of natural fractals are weather (think of a satellite image of a hurricane and television footage of a swirling tornado), flow of fluids in a stream, river and even our bodies, geological activity like earthquakes, the orbit of a planet, music, behavior of groups of animals and even economic changes in a country.

The colorful image of the fractal can be used to model how living things survive in whatever environment

Fractals and Jackson Pollock

Early in his career as a painter, the American artist Jackson Pollock struggled to find a way to express his artistry on canvas. Ultimately, he unlocked his creativity by dripping house paint onto huge canvasses using a variety of objects including old and hardened paintbrushes and sticks. The result was a visual riot of swirling colors, drips, splotches, and cross-canvas streaks.

There was more to Pollock's magic than just the random flinging of paint onto the canvas. Typically, he would begin a painting by using fluid stokes to draw a series of looping shapes. When the paint dried, Pollock often connected the shapes by using a slashing motion above the canvas. Then, more and more layers of paint would be dripped, poured and hurled to create an amazing and colorful spider-web of trails all over the huge canvas.

Pollock's paintings are on display at several of the world's major museums of modern art, including the Museum of Modern Art in New York and the Guggenheim Museum in Venice, Italy, and continue to amaze many people. The patterns of paint actually traced Pollock's path back and forth and around the canvas as he constructed his images. One reason that these patterns have such appeal may be because of their fractal nature.

In 1997, physicist and artist Richard Taylor of the University of New South Wales in Australia photographed the Pollock painting *Blue Poles, Number 11, 1952*, scanned the image to convert the visual information to a digital form, and then analyzed the patterns in the painting. Taylor and his colleagues discovered that Pollock's artistry represented fractals. Shapes or patterns of different sizes repeated themselves throughout the painting. The researchers postulated that the fact that fractals are so prevalent in the natural world makes a fractal image pleasing to a person at a subconscious level.

Analysis of Pollock while he was painting and of paintings over a 12-year period from 1943–1952 showed that he refined his construction of fractals. Large fractal patterns were created as he moved around the edge of the canvas, while smaller fractal patterns were produced by the dripping of paint onto the canvas.

Pollock died in a high-speed car crash in 1956, long before the discovery of fractals that powered his genius.

they are in. The complexity of a fractal mirrors the complexity of nature. The rigid rules that govern fractal formation are also mirrored in the natural world, where the process of constant change that is evolution takes place in reasonable way. If a change is unreasonable, such as the sudden appearance of a strange mutation, the chance that the change will persist is remote. Fractals and unreasonable changes are not compatible.

Let us consider the fractal modeling of a natural situation. An example could be the fate of a species of squirrel in a wooded ecosystem that is undergoing a change, such as commercial development. The squirrel's survival depends on the presence of the woods. In the fractal model, the woods would be colored black and would be the central image of the developing fractal. Other environments that adversely affect the squirrel, such as smoggy air or the presence of acid rain, are represented by different colors. The colors indicate how long the squirrel can survive in the adverse condition. For example, a red color might indicate a shorter survival time than a blue color. When these conditions are put together in a particular mathematical equation, the pattern of colors in the resulting fractal, and the changing pattern of the fractal's shape, can be interpreted to help predict how environmental changes in the forest will affect the squirrel, especially at the border of the central black shape, where the black color meets the other colors in the image.

MODELING HURRICANES AND TORNADOES

Nonliving systems such as hurricanes and tornadoes can also be modeled this way. Indeed, anything whose survival depends on its surroundings is a candidate for fractal modeling. For example, a hurricane draws its sometimes-terrifying strength from the surrounding air and sea. If the calm atmosphere bordering a hurricane, and even the nice sunny weather thousands of miles away could be removed somehow, the hurricane would very soon disappear.

NONLIVING SYSTEMS

Other nonliving systems that can be modeled using fractals include soil erosion, the flicking of a flame and

the tumbling or turbulent flow of a fluid like water. The movement of fluid through the tiny openings in rocks is another example. Indeed, oil companies use fractal modeling to try to unravel the movement of oil through rock formations to figure out where the best spot to drill might be to get the most oil with the least expense and danger.

ASTRONOMY

Fractals can be useful in understanding the behavior of events far from Earth. Evidence is mounting that the arrangement of galaxies in the inky vastness of space is fractal-like, in that the galaxies are somewhat similar in shape and are clustered together in a somewhat ordered way. "Clustered together" is relative; the galaxies are millions of light years apart. Still, in the infinity of space, the galaxies can be considered close neighbors. While the fractal nature of the universe is still controversial, it does make sense, because here on Earth the natural world beats to a fractal rhythm.

CELL PHONE AND RADIO ANTENNA

Fractals also have real-world applications in mechanical systems. One example is the design of the antennas that snag radio and other waves that pass through the air. A good antenna needs a lot of wave-trapping wire surface. Having a long and thin wire is not the best design. But, because some antennas need to fit into a narrow space (think of the retractable antenna on a car and in a cellular phone), there is not much room for the wire. The solution is fractals, whose mix of randomness (portions of the fractal) and order (the entire fractal) can pack a greater quantities of material into a smaller space.

By bending wires into the multi-star-shaped fractal that is the star-shaped Koch's curve, much more wire can be packed into the narrow confines of the antenna barrel. As an added benefit, the jagged shape of the snowflake-shaped fractal actually increases the electrical efficiency of the antenna, doing away with the need to have extra mechanical bits to boost the antenna's signal-grabbing power. Some companies use fractal antennas in cell phones. This innovation has proven to be more efficient than the traditional straight piece of wire antennas, they are cheaper to make, and they can be built right into the body of the phone, eliminating the pull-up antenna. The next time your cell phone chirps, the incoming connection might be due to a fractal.

COMPUTER SCIENCE

Another use of fractals has to do with computer science. Images are compressed for transmission as an email attachment in various ways such as in JPEG or GIF formats. A route of compression called fractal compression, however, enables the information in the image to be squeezed into a smaller, more easily transmitted bundle at one end, and to be greatly enlarged with a minimal loss of image quality.

There are many fractal equations that can be written, and so there are many images of fractals. The images are often beautiful; many sites on the Internet contain stunning fractal images available for download.

Where to Learn More

Books

Barnsley, M.F. *Fractals Everywhere*. San Francisco: Morgan Kaufmann, 2000.

Lesmoir-Gordon, N., W. Rood, R. Edney, and R. Appignanesi. *Introducing Fractal Geometry*. New York: National Book Network, 2000.

Mumford, D., C. Series, and D. Wright. *Indra's Pearls: The Vision of Felix Klein*. Cambridge: Cambridge University Press, 2002.

Web sites

Connors, M.A. "Exploring Fractals." *University of Massachusetts Amherst* <http://www.math.umass.edu/~mconnors/fractal/fractal.html> (September 8, 2004).

Lanius, Cynthia. "Why Study Fractals?" *Rice University School Math Project* <http://math.rice.edu/~lanius/fractals/WHY/> (September 8, 2004).

Fractions

Overview

A fraction is a number written as two numbers with a horizontal or slanted line between them. The value of the fraction is found by dividing the number above the line by the number below the line. Not only are fractions a basic tool for handling numbers in mathematics, they are used in daily life to measure and price objects and materials that do not come in neatly countable, indivisible units. (We must often deal with a fraction of a pizza or a fraction of an inch, but we rarely have to deal with a fraction of an egg.) Fractions are closely related to percentages.

Fundamental Mathematical Concepts and Terms

WHAT IS A FRACTION?

Every fraction has three parts: a horizontal or slanted line, a number above the line, and a number below the line. The number above the line is the "numerator" and the number below the line is the "denominator." For example, in the fraction 3/4 (also written ¾), the numerator is 3 and the denominator is 4.

The fraction 3/4 is one way of writing "3 divided by 4." In general, a fraction with some number a in the numerator and some number b in the denominator, a/b, means simply "a divided by b." For example, writing 4/2 is the same as writing $4 \div 2$. Because division by 0 is never allowed, a fraction with 0 in the denominator has no meaning.

You can think of a fraction as a way to say how many portions. For example, if you slice 1 pizza into 8 equal-sized parts, each piece is an eighth of a pizza, 1/8 of a pizza. If you put 3 of these pieces on your plate, you have three eighths of the pizza, or 3/8.

TYPES OF FRACTIONS

There are different kinds of fractions. A proper fraction is a fraction whose value is less than 1, and an improper fraction is a fraction whose value is greater than or equal to 1. For example, 3/5 is a proper fraction, but 5/3 is an improper fraction. Despite the disapproving sound of the word "improper," there is nothing mathematically wrong with an improper fraction. The only difference is that an improper fraction can be written as the sum of a whole number and a proper fraction: 5/3, for example, can be written as $1 + 2/3$.

A unit fraction is any fraction with 1 in the numerator. This kind of fraction is so common that the English language has special words for the most familiar ones: 1/2 is a "half," 1/3 is a "third," and 1/4 is a "quarter."

Gas prices for (from top) plus, premium, and diesel, are typically shown with fractions denoting tenths of a cent.
AP/WIDE WORLD PHOTOS. REPRODUCED BY PERMISSION.

Two or more fractions are called equivalent if they stand for the same number. For example, 4/2 and 8/4 are equivalent because they both equal 2.

A lowest-terms fraction is a fraction with all common terms canceled out of the numerator and denominator. A "common term" of two numbers is a number that divides evenly into both of them: 2 is a common term of 4 and 16 because it goes twice into 4, eight times into 16. For the fraction 2/16, therefore, 2 is a common term of both the numerator and denominator, and so the fraction 2/16 is not a lowest-terms fraction. We can make 2/16 into a lowest-terms fraction by dividing the numerator and the denominator by 2.

A mixed fraction is made up of an integer plus a fraction, like 1 + 1/2. In cooking and carpentry (but never in mathematics), a mixed fraction is written without the "+" sign: 1 1/2.

RULES FOR HANDLING FRACTIONS

To be useful, fractions must be added, subtracted, multiplied, and divided by other numbers. The rules for how to do each of these things are given in Table 1.

FRACTIONS AND DECIMALS

Fractions are closely related to another mathematical tool used in science, business, medicine, and everyday life, namely decimal numbers. A number in decimal form, such as 3.1415, is shorthand for a sum of fractions: each of the numbers to the right of the decimal point (the "." in 3.1415) stands for a fraction with a multiple of 10 in its denominator. The first position to the right of the decimal point is a tenth, the second is a hundredth, the third is a thousandth, and so forth: .1 = 1/10, .01 = 1/100, .001 = 1/1,000, and so on. Therefore we can write any decimal number as a sum of fractions; for example, 3.1 = 3 + (1/10).

FRACTIONS AND PERCENTAGES

Fractions are also close cousins of percentages, which are fractions with 100 in the denominator. For example, to say "50 percent" is exactly the same as saying "fifty hundredths" (50/100). This fraction, 50/100, can be reduced to a least-terms fraction by dividing the numerator and the denominator by 50 to get 1/2. Accordingly, "50 percent" is the same as "half."

However, if percentages are just fractions, why use percentages? We do so because they give us a quick, useful way of relating one thing (a count or concept) to another. Say, for example, that we want to know how many people in a population of 150 million are unmarried. We conduct a survey and find out that the answer is 77 million. To describe this fact by reeling off the raw data—"77 million out of 150 million people in this population are unmarried"—would be truthful but clumsy. We can make things a little better by writing the two numbers as a fraction, 77,000,000/150,000,000, and then converting this into a least-terms fraction by dividing the numerator and denominator by 1,000,000. This gives us 77/150, which is more compact than 77,000,000/150,000,000, but is still hard to picture in the mind: how much is 77/150? Most of us have to do a little mental arithmetic to even say whether 77 is more than half of 150 or not. (It's a little more.) The handiest way to express our results would be to use a fraction with a familiar, easy-to-handle denominator like 100—a percentage. One way to do this is to divide 77 by 150 on a calculator, read off the answer in decimal form as .5133333 (the 3s actually go on forever, but the calculator cannot show this), and round off this number to the nearest hundredth. Then we can say, "51 percent of this population is married"—51/100.

By rounding off, however, we throw away a little information. (If all you keep from .5133333 is .51, the .0033333 is gone—lost.) In this case, however, as in great many real-life cases, the loss is not enough to matter. It is small because a hundredth is a small fraction. If we rounded off to thirds instead of hundredths, we would lose much more information: the closest we could come

Operation	Rule	Example
Multiply a fraction by an integer, *n*	$n\frac{a}{b} = \frac{na}{b}$	$4 \times \frac{3}{5} = \frac{4 \times 3}{5} = \frac{12}{5}$
Divide a fraction by an integer, *n*	$\frac{a}{b} \div n = \frac{a}{bn}$	$\frac{1}{2} \div 4 = \frac{1}{2 \times 4} = \frac{1}{8}$
Multiply fractions	$\frac{a}{b}\frac{c}{d} = \frac{ac}{bd}$	$\frac{1}{3} \times \frac{4}{5} = \frac{1 \times 4}{3 \times 5} = \frac{4}{15}$
Divide fractions	$\frac{a}{b} \div \frac{c}{d} = \frac{a}{b}\frac{d}{c} = \frac{ad}{bc}$	$\frac{1}{3} \div \frac{4}{5} = \frac{1}{3} \times \frac{5}{4} = \frac{5}{12}$
Add fractions	$\frac{a}{b} + \frac{c}{d} = \frac{ad + bc}{bd}$	$\frac{2}{9} + \frac{1}{7} = \frac{2 \times 7 + 1 \times 9}{9 \times 7} = \frac{23}{63}$
Subtract fractions	$\frac{a}{b} - \frac{c}{d} = \frac{ad - bc}{bd}$	$\frac{2}{9} + \frac{1}{7} = \frac{2 \times 7 - 1 \times 9}{9 \times 7} = \frac{5}{63}$

Table 1: Rules for handling fractions.

to 77/150 would be 2/3, which is 66%, which is much farther from the truth than 51% is. By expressing information as percentages (in numbers of hundredths), we get three advantages: (1) accuracy, because hundredths allow for pretty good resolution or detail; (2) compactness, because a percentage is usually easier to write down than the raw numbers; and (3) familiarity—because we are used to them.

ALGEBRA

All the rules that apply to adding, subtracting, multiplying, and dividing fractions are used constantly in algebra and higher mathematics. Simple fractions have only an integer in the numerator and an integer in the denominator, but there is no reason not to put more complicated mathematical expressions in the numerator and denominator—and we often do. For example, we can write expressions such as $x^2 / (9 + x^2)$ where x stands for an unknown number. Any material above the line in fraction-like expression such as 1/2 or to the left side of a fraction written in linear form such as 1/2), no matter how complicated it is, is the "numerator," and any material below the line is the "denominator." Such expressions are added, subtracted, multiplied, and divided using exactly the same rules that apply to ordinary number fractions like 3/4.

A Brief History of Discovery and Development

Fractions were invented about 4,000 years ago so that traders could keep track of how they were dividing goods and profits. For instance, if three traders own a grain-selling stall together, and 10 sacks of grain are sold at 1 denarius apiece, how much profit should the books record for each owner? Three does not go evenly into 10, so we have a fraction, 10/3. This is an improper fraction, and can be reduced to 3 + 1/3. Each trader should get 3 whole denarii and credit on the books for 1/3 more.

Many ancient civilizations developed some form of dealing with fractions—the Mayan, the Chinese, the Babylonian, the Egyptian, the Greeks, the Indians (in Asia), and the Romans. However, for centuries these systems of writing and dealing with fractions had severe limits. For example, the Egyptians wrote every fraction as a sum of unit fractions (fractions with a "1" in the numerator). Instead of writing 2/7, the Egyptians wrote their symbols for 1/4 + 1/28 (which, if you do the math, does add up to 2/7). The Romans did not write down fractions using numbers at all, but had a limited family of fractions that they referred to by name, just as we speak of a half, a third, or a quarter. All the other ancient systems had their own problems; in all of them it was very difficult to do

calculations with fractions, like adding and subtracting and multiplying them. Finally, around 500 A.D., a system of number-writing was developed in India that was similar to the one we use now. In fact, our system is descended from that one through Arab mathematics.

Fraction theory advanced in the 1600s with the development of practical applications for continued fractions. Continued fractions are fractions that have fractions in their denominators, which have fractions in their denominators, and so on, forever if need be.

Continued fractions were originally used for designing gears for clocks and other mechanisms. Today they are used in the branch of mathematics called "number theory," which is used in cryptography (secret coding), computer design, and other fields.

Real-life Applications

COOKING AND BAKING

Fractions are basic to cooking and baking. Look at any set of cup measures or spoon measures: they are all marked in fractions of a cup (or, in Europe, fractions of a liter). A typical cup-measure set contains measures for 1 cup, 1/2 cup, 1/3 cup, 1/4 cup, and 1/8 cup; a typical spoon set contains measures for a tablespoon, a teaspoon, 1/2 teaspoon, 1/4 teaspoon, 1/8 teaspoon; some also include 1/2 tablespoon and 1 1/2 tablespoon.

Not only are measurements in cooking and baking done in fractions, a cook must often know at least how to add and multiply fractions in order to use a recipe. Recipes might only given for a single batch: if you want to make a half batch, or a double or triple batch, you must halve, double, or triple all the fractional measurements in the recipe. Say, for example, that a cookie recipe calls for 2 2/3 cups of flour and you want to make a triple batch. How much flour do you need to measure?

There are several ways to do the math, but all require a knowledge of fractions. One way is to write the mixed number 2 2/3 as a fraction by first noting that 2 = 6/3. Therefore, by the rule for adding fractions that have the same denominator, 2 2/3 = 6/3 + 2/3 = 8/3. To triple the amount of flour in the batch, then, you multiply 8/3 cups by 3:

$$\frac{8}{3} \times 3 = \frac{8 \times 3}{3} = \frac{24}{3}$$

At this point you can either get out your 1/3-cup measure and measure 24 times, which is a lot of work, or you can try reducing 24/3 to a mixed fraction. If you try reducing the fraction, you will probably discover at once that 24 / 3 = 8. Therefore, you can measure eight times with your 1-cup measure and move on to the next ingredient on the list.

RADIOACTIVE WASTE

A continuing political issue in nuclear capable countries is the question of what to do with nuclear waste. Such waste is an unwanted by-product of making electricity from metals like uranium and plutonium. Nuclear waste gives off radiation, a mixture of fast-moving atomic particles and invisible, harmful kinds of light that at low levels may cause cancer and at high levels can kill living things. Only over very long periods will radioactive waste slowly become harmless as it breaks down naturally into other elements. This happens quickly for some substances in the waste mixture, slowly for others. How quickly a substance loses its radioactivity is expressed as a fraction, the "half-life" of the substance. The half-life of a substance is the time it takes for any fixed amount of the substance to lose 1/2 of its radioactivity. For the element plutonium, which is found in most nuclear waste, the half-life is about 24,000 years. That is, no matter how much plutonium you start out with at time zero, after 24,000 years you will have half as much plutonium left. (But not half as much radioactivity, exactly, since some of the elements that plutonium breaks down into are radioactive themselves, with half-lives of their own, and must break down further before they can become harmless.)

By multiplying fractions, it is possible to answer some questions about how much radioactive waste will remain after a certain time. For instance, after two half-lives, how much of 1 kilogram (kg) of plutonium will be left? This is the same as asking what is a half of a half, which is the same as multiplying 1/2 times itself: 1 kg × 1/2 × 1/2 = 1/4 kg.

This can be carried on for as many steps as we like. For example, how much of 1 kg of plutonium will be left after 10 half-lives (that is, after 240,000 years)? The answer is 1 kg × 1/2 × 1/2 × 1/2 × 1/2 × 1/2 × 1/2 × 1/2 × 1/2 × 1/2 × 1/2 = 1 / 1024 kg.

This shows that the plutonium will never completely disappear. The denominator gets larger and larger, which makes the value of the fraction smaller but cannot make it equal to zero.

MUSIC

Fractions and rhythm Fractions are used throughout music. In Western music notation, the time-values of notes are named after fractions: besides the whole note, which lasts one full beat, there is the half-note, which lasts

only half a beat, and the quarter note, eighth note, sixteenth note, and so forth. Notice that these fractions—1/2, 1/4, 1/8, 1/16—all have multiples of 2 in the denominator. In fact, each fraction in the series is the previous fraction times 1/2. That is, each standard type of note lasts 1/2 as long as the next-longest type. Music notation also has "rest" symbols, marks that tell you how long to be silent. Just as there are notes with various values, there are rest symbols with various time values—whole, half, quarter, and eighth rests.

Nor are we limited to the beat fractions given above. Each of the standard notes can also be marked with a dot, which indicates that the duration of the note is to be increased by 1/2. This is the same as multiplying the time value of the original note by 3/2. So, for example, the time value of a dotted eighth note is given by $1/8 \times 3/2$, which is 3/16. And by tying three notes together with an arc-shaped mark (a "tie") and writing the number "3" by the arc, we can show that the musician should play a "triplet," a set of three notes in which each note lasts 1/3 of a beat.

Fractions and the musical scale A single guitar string can produce many different notes. The guitar player pushes the string down with a fingertip on a steel bar called a "fret," shortening the part of the string that vibrates freely. The shorter the freely vibrating part of the string, the higher the note. The Greeks also made music using stringed instruments, and they noticed several thousand years ago that the notes of their musical scale—the particular notes that just happen to be pleasing to the human ear—were produced by shortening a string to certain definite fractions of its full length. Sounding the open string produced the lowest note: the next-highest pleasing note was produced by shortening the string to 4/5 of its open length. Shortening the string to 3/4, 2/3, and 3/5 length—each fraction smaller than the last, each note higher—produced the three notes in the Greek 5-note scale. Shortening a string to 1/2 its original length produces a note twice as high as the open string does: musically, this is considered the same note, and the scale starts again.

Modern music systems have more than 5 notes; in the Western world we use 12 evenly-spaced notes called "semitones." Seven of these notes have letter names—A, B, C, D, E, F, and G—and five are named by adding the terms "sharp" or "flat" to the letters. These musical choices are built right into our instruments. If you look at the neck of a guitar, for example, you will see that the frets divide it up into 12 parts. Why 12? The ancients decided to see what would happen if they divided the fractional string lengths of the Greeks' 5-note scale into similar fractions. That is, if one pleasing note is produced by shortening the string to 2/3 its open length, what note do we get if we shorten the string to 2/3 of that shorter length? The vibrating part of the string is then $2/3 \times 2/3 = 4/9$ the length of the open string. But this is less than half the length of the string, making the note an octave too high, so we double the fraction to lengthen the string and lower the note: $4/9 \times 2 = 8/9$. And indeed, the fret for playing a B on the A string of a guitar does shorten the string to 8/9 of its open length. By similarly multiplying the fractions that gave the other notes in the original Greek scale, people discovered 12 notes—the semitones we use today. Later, in the 1600s, people decided that they would space the notes slightly differently, based on multiples of the 12th root of 2 rather than on fractions. This makes it easier for instruments to be tuned to play together in groups, as the notes are spaced perfectly evenly, and as long two instruments match on one semitone they will match on all the others too. These modern notes are close to the fraction-based notes, but not exactly the same.

SIMPLE PROBABILITIES

Many U.S. states make money through lotteries, public games in which any adult can buy one or more tickets. The money spent on tickets is pooled, the state keeps a cut, and the rest is given to a single winning ticket-buyer who is chosen by chance. Some states have become dependent on the money they make from the lotteries, which now totals many billions of dollars every year. If you bought a lottery ticket, what would your chances of winning? Mathematically, we would ask: what is the probability that you will win?

A "probability" is always a number between 0 and 1. Zero is the probability of an event that can't possibly happen; 1 is the probability of an event that is sure to happen; and any number between 0 and 1 can stand for the probability of an event that might happen. If you buy one lottery ticket in which, say, 10 million other people have bought a ticket, then the probability that you will win is a unit fraction with 10 million in the denominator: a 1-ticket chance of winning = 1/10,000,000.

If you buy two tickets, your chance of winning is this fraction multiplied by 2: a 2-ticket chance of winning = 2/10,000,000. This fraction can be reduced to a lowest-terms fraction by dividing both the numerator and denominator by 2 to yield 1/500,000. Accordingly, buying two tickets doubles your chances of winning. On the other hand, double a very small chance is still a very small chance.

Lottery chances are typical of a certain kind of probability encountered often in everyday life, namely, when some number of events is possible (say, N events), but only one of these N events can actually happen. If all

N events are equally likely, then the chance or probability of any one of them happening is simply the fraction 1/*N*.

The math of probabilities gets much more complicated than this, but simple fractional probabilities can be important in daily life. Consider, for instance, two people who are considering having a baby. Many serious diseases are inherited through defective genes. Each baby has two copies of every gene (the molecular code for producing a certain protein in the body), one from each parent, and there are two kinds of defective genes, "dominant" and "recessive." For a disease controlled by a dominant defect, if the baby has just 1 copy of the defective gene from either parent, it will have the disease. If one parent carries one copy of the dominant defective gene in each of their cells, the probability that the baby will have the dominant gene (and therefore the disease) is 1/2; if both parents have one copy of the dominant gene, the probability that the baby will have the dominant gene is 3/4. Parents who are aware that they carry defective genes cannot make informed choices about whether to have children or not unless than can understand these fractions (and similar ones).

Key Terms

Equivalent fractions: Two fractions are equivalent if they stand for the same number (that is, if they are equal). The fractions 1/2 and 2/4 are equivalent.

Improper fraction: A fraction whose value is greater than or equal to 1.

Least-terms fraction: A fraction whose numerator and denominator do not have any factors in common. The fraction 2/3 is a least-terms fraction; the fraction 8/16 is not.

Proper fraction: A fraction whose value is less than 1.

Unit fraction: A fraction with 1 in the numerator.

OVERTIME PAY

In many jobs, workers who put in hours over a certain agreed-only weekly limit—"overtime"—get paid "time and a half." This means that they are paid at 3/2 or 1 + 1/2 times their usual hourly rate.

Multiplying this fraction by your usual hourly rate gives the amount of money your employer owes you for your overtime.

TOOLS AND CONSTRUCTION

Most of us have to use tools at some time or another, and millions of people make a living using tools. An understanding of fractions is necessary to do any toolwork much more complicated than hammering in a nail. To begin with, all measurements, both in metal and wood, are done using rulers or measuring tapes that are divided into fractions of an inch (in the United States) or of a meter (in Europe). The fractions used are based on halving: if the basic unit of measure is an inch, and then the ruler or tape is marked at 1/2 inch, 1/4 inch, 1/8 inch, and 1/16 inch, each fraction of being half as large as the next-largest one. So to read a ruler or a tape measure it is necessary to at least be able to read off the fractions. Further, in making anything complex—framing a house, for example—it is necessary to be able to add and subtract fractions. For example, you are framing a wall that is 6 feet (72 inches) wide. You have laid down "two-by-fours" at the ends of the wall, at right angles to it where the other walls meet, like the upright arms of a square "U." (A two-by-four was 2 inches thick and 4 inches wide many years ago, but today is 1 1/2 inches thick and 3 1/2 inches wide. Notice that in carpentry, as in cooking, it is acceptable to write "1 1/2" for 1 + 1/2.) Now you want to cut a two-by-four to lay down along the base of the 6-foot wall, in the space that is left by the two two-by-fours that are already down at right angles: you want to put in the bottom of the square "U." How long must it be?

You could, in this case, just measure the distance with a tape measure. But there are many occasions, in building a house, when it is simply not possible to measure a distance directly, and we'll pretend that this is one of them (because it's relatively simple). Each of the two-by-fours uses up 3 1/2 inches of the 72 inches of wall. There are several ways to do the problem: one is to add 3 1/2 + 3 1/2 to find that the two two-by-fours use up 7 inches of space. Since 72 − 7 = 65, you want to cut a board 65 inches long.

Another place where fractions pop up in the world of tools and construction is in dimensions of common tools. United States drill bits, for instance, typically come in widths of the following fractions of an inch: 1/4, 3/16, 5/32, 1/8, 7/64, 3/32, 5/64, and 1/16.

FRACTIONS AND VOTING

Simple fractions like and 1/4, 1/3, and 2/3 have a common-sense appeal that leads us to use them again and again in everyday life. They appear often in politics, for example. The United States Constitution states that the President (or anyone else who could be impeached) can

only be convicted if 2/3 of the members of the Senate who are present agree. Some other fraction could have been used—4/7, say—but would not have been as simple.

One of the most famous fractions in political history, 3/5, appears in the United States Constitution, Article I, Section 2, which reads as follows: "Representatives and direct Taxes shall be apportioned among the several States which may be included within this Union, according to their respective Numbers, which shall be determined by adding to the whole Number of free Persons, including those bound to Service for a Term of Years, and excluding Indians not taxed, three fifths of all other Persons." Translated into plain speech, this means that the more people live in a state, the more congresspeople would be needed to represent it in the House of Representatives, giving it more voting power. The phrase "all other Persons" was an indirect reference to "slaves." Because of this clause in the Constitution, slaves, though they had no human rights, would count toward allotting congresspeople (and thus political power) to Southern states. The Southern states wanted the Constitution to count slaves as equal to free persons for the purposes of allotting state power in Congress, and the Northern states wanted slaves counted as a smaller fraction or not at all; James Madison proposed the fraction 3/5 as a compromise. The rule was ultimately canceled by the Fourteenth Amendment after the Civil War, but it did play an important part in U.S. history: Thomas Jefferson was elected to the Presidency in 1800 by Electoral College votes of Southern states derived from the three-fifths rule. (By the way, neither the original Constitution or the 14th Amendment counted women at all: you might say that they were counted at 0/5 until the 19th Amendment gave them the legal right to vote in 1920.)

In 2004, a bill was proposed to give teenagers fractional voting rights in California. If the bill had passed, 14- and 15-year-olds would have been given votes worth 1/4 as much as those of adults and 16- and 17-year-olds would have been given votes worth 1/2 as much as those of adults. (All people 18 years and older already have the right to vote, each counted as one full vote.) The intent was to teach teenagers to take the idea of participating in democracy seriously from a younger age. Some European countries such as the United Kingdom have seriously considered lowering the voting age to 16—with no fractions involved.

Where to Learn More

Web sites

"Egyptian Fractions." Ron Knott, Dept. of Mathematics and Statistics, University of Surrey, Guildford, UK, October 13, 2004. <http://www.mcs.surrey.ac.uk/Personal/R.Knott/Fractions/egyptian.html#ef> (April 6, 2005).

"Causes of Sea Level Rise." Columbia University, 2005. <http://www.columbia.edu/~epg40/elissa/webpages/Causes_of_Sea_Level_Rise.html> (April 4, 2005).

"Fractions." Math League Multimedia, 2001. <http://www.mathleague.com/help/fractions/fractions.htm#whatisafraction> (April 6, 2005).

"Fraction." Mathworld, Wolfram Research, 2005. <http://mathworld.wolfram.com/Fraction.html> (April 14, 2005).

Functions

Overview

A function is a rule for relating two or more sets of numbers. The word "function" is from the Latin for perform or execute, and was first used by the German mathematician Gottfried Wilhelm von Leibniz, one of the inventors of calculus, in the late 1600s. A function is often written as an equation, that is, as two groups of mathematical symbols separated by an equals sign. A function can also be expressed as a list of numbers on paper or in computer memory.

A function in mathematics is, more or less, what the sentence is in English: it expresses a complete mathematical thought. Much of mathematics consists of functions and the rules that relate them to each other. To use mathematics in the real world means looking for functions to describe how things behave, and then using those functions to design, understand, predict, or control the things they describe.

Fundamental Mathematical Concepts and Terms

Mathematicians divide all known kinds of function into types. Each type of function has different properties and different uses. Some of the most common types of function are polynomials, exponentials, and trigonometric functions. Lists of more complicated, unusual functions are published as lists of "special functions."

FUNCTIONS AND RELATIONS

A function is a particular type of "relation." A relation is a set of ordered pairs of numbers. For example, the three pairs (2,1), (4,2), and (3,1) give a relation. It is usual in mathematics to refer to each left-hand number in a pair as an x and to each right-hand number as a y. In this relation, x can be 2, 4, or 3, and y can be 1 or 2.

A relation is also a function if each y goes with only one x. The relation (2,1), (4,2), and (3,1) is not a function because a y value of 1 goes with two values of x, namely 2 and 3. For a relation to be a function, each x must be paired with one and only one x.

In practice, the word "function" is often used to include equations that assign more than one y value to each x value. Such equations are sometimes called "multiple-valued functions" and are also useful in dealing with the real world.

HOW FUNCTIONS ARE DESCRIBED

Functions are usually written as an equations. This is done because most functions relate not only a few pairs of

numbers, as in the example already given, but many pairs of numbers—too many to write down. For example, if we decide to use x and y to stand for any of the positive counting numbers (1, 2, 3, . . .), then the equation $y = 2x$ describes a function that relates an infinite number of x's to an infinite number of y's as shown in Table 1.

The only practical way to write down such a function is in the form of an equation.

Can such a simple function as $y = 2x$ describe something in the real world? Certainly. If you are climbing a steep mountain that goes up two feet for every foot you go forward, then the vertical (upward) distance you travel is twice as great as the horizontal (sideways) distance you travel. If we use y to stand for the distance you have traveled vertically and x to stand for the distance you have traveled horizontally, then $y = 2x$.

The letters x and y are called "variables" because they can vary, that is, take on different values. The numbers that x can stand for are called the function's "domain" and the numbers that y can stand for are called its "range." In the example above, the domain and the range both consist of all the positive counting numbers, 1, 2, 3, and so forth. Many symbols other than x and y are used to stand for variables; these two are merely the most common, by tradition.

A function can be thought of as a sort of number machine that takes in an x value and puts out a y value. The y value can thus be thought of as depending on the x value, so x is sometimes called the "independent variable" and y the "dependent variable." If we are using a function to describe cause-and-effect in the real world, we describe the cause as an independent variable and the effect as a dependent variable. For example, if a rocket is pushing on a spacecraft with a certain force, and we want to write a function that relates this force to the acceleration (increase of speed) of the spacecraft, we write the force as the input or independent variable and the acceleration of the spacecraft as the output or dependent variable. It is the acceleration that depends on the force, not the other way around.

There are also functions that have more than one independent variable. We put two (or more) numbers into the function rather than one, and the function produces a single number as output.

y's	*x's*
2 = 2 ×	**1**
4 = 2 ×	**2**
6 = 2 ×	**3** ... and so on, forever

Table 1. The function $y = 2x$, where x can be any positive counting number.

Real-life Applications

MAKING AIRPLANES FLY

One of the basic laws of physics is that energy can neither be created nor destroyed. It can, however, be moved from one form to another. For example, a fluid (gas or liquid) can store energy in several ways: as heat, as motion, or as pressure. A function called Bernoulli's equation, named after its discoverer, Swiss mathematician Daniel Bernoulli (1700–1782), describes how a moving fluid can move its energy between pressure and motion. According to Bernoulli's equation, the faster a fluid flows, the lower its pressure.

Bernoulli's equation—or, rather, the physical effect described by it—is what keeps airplanes up. An airplane wing is shaped so that as it slices through the air, the air that flows over the rounded top of the wing has to travel farther than air under the flat bottom of the wing. The air on top and on bottom must make the trip from the leading edge of the wing to the trailing edge in the same amount of time, so the wing on the upper surface, which has farther to go, is forced to flow faster. But this means, by Bernoulli's equation, that its pressure is decreased. As a result, there is more pressure—more force—per square foot on the wing's bottom than on its top. This difference in pressure is what holds the plane up. In the case of a Boeing 747, the pressure difference between the top and bottom sides of the wing is over 100 pounds per square foot.

You can test the Bernoulli principle by holding a sheet of paper by one edge so that it droops away from you, and blowing on the top of it. The paper will rise into the jet of air from your mouth because the pressure in the moving air is lower.

Bernoulli's equation only holds true for flight slower than the speed of sound (about 741 mph [1,193 km/h], at sea level). At speeds faster than the speed of sound—supersonic flight—the behavior of air is so different that functions other than Bernoulli's equation must be used to design aircraft wings.

GUILLOCHÉ PATTERNS

Look at any piece of paper money from almost anywhere in the world. Somewhere on the bill—on United States currency, it is around the border—you will see a dense, complex pattern of curving, intersecting lines. This

design is known as a Guilloché pattern and they have been used since the 1850s on paper money, stocks, bonds, and other official documents in order to make them more difficult to counterfeit. Today they are also used on laminated plastic cards such as some identification cards and driver's licenses.

Guilloché patterns were originally produced by mechanical means. Today they are produced mathematically on computers, using the functions called sinusoids. Sinusoids are functions that look, when graphed, like smooth, wavy lines. A Guilloché pattern is produced by multiplying sinusoids, shifting them, adding them, and placing one on top of another. (Straight lines may also be added to make the pattern even more complex.)

Modern money is protected from counterfeiting more by invisible inks, hidden plastic threads, and other hard-to-copy features than by its Guilloché patterns, but a finely-printed Guilloché pattern still makes it more difficult to produce a convincing fake of a piece of money or an identification card. A number of software packages are for sale that make Guilloché patterns.

THE MILLION-DOLLAR HYPOTHESIS

In 1859, the same year that saw the publication of Charles Darwin's theory of evolution in *The Origin of Species*, the German mathematician Georg Friedrich Bernhard Riemann (1882–1866) made a guess about a special function, the "zeta function." His guess, or hypothesis, was that all the zeroes of this function—all the values of the independent variable for which the function equals 0—lie on a certain line.

Mathematicians have been trying to prove the Riemann hypothesis for well over century. So important is the Riemann hypothesis that in 2000, the Clay Mathematics Institute of Cambridge, Massachusetts, announced that it would give $1,000,000 to the first person to prove it. Computers have been used to test millions of zeroes of the zeta function and found that all of them, so far, do lie on the line that Riemann described. But the zeta function has an infinite number of zeroes, so no computer study can prove that *all* the zeroes lie on the line.

The Riemann hypothesis matters to real-world mathematics because it is one of the most important ideas in the study of prime numbers. A prime number is any whole number that cannot be evenly divided by any number smaller than itself except 1. (The first 10 prime numbers are 2, 3, 5, 7, 11, 13, 17, 19, 23, and 29.) Large prime numbers (primes dozens or hundreds of digits long) are at the heart of modern cryptography, the science of sending secret messages—and cryptography is at the heart of our modern economy. Cryptography allows banks and Internet users to send money, credit-card numbers, and other private information as coded messages without fear that thieves will be able to read or fake those messages. If the Riemann hypothesis concerning the zeta function is proved, it will become easier to discover new large primes.

In June 2004, a French mathematician working at Purdue University, Louis de Branges de Bourcia (1932–) claimed to have discovered a proof of the Riemann hypothesis. His proof was long and complicated, and mathematicians could not quickly agree whether he had solved the puzzle or not. As of late 2005, de Bourcia's claim had not been proved or disproved by other mathematicians.

FINITE-ELEMENT MODELS

A computer can predict the weather—more or less—by "modeling" it, that is, by working with a three-dimensional mesh or net of numbers. Each separate hole in the mesh or net—also called an element of the mesh—stands for a small piece of the atmosphere. Numbers attached to each element represent the pressure, temperature, motion, and other properties of that piece of air. Next, functions are programmed into the computer to describe the physical laws that the element must obey. Foremost among these functions are the Navier-Stokes equations, which relate air pressure to speed and acceleration. This set of equations is named after French physicist Claude-Louis Navier (1785–1836) and Irish mathematician Gabriel Stokes (1819–1903), the men who discovered them.

The computer feeds the numbers on the mesh into the functions that describe what should happen to them next. As output, the functions produce a new set of numbers that is, in effect, a prediction of what the weather will be a little bit into the future. This computation is done repeatedly, until eventually a prediction for the weather hours, days, or weeks into the future is produced. Will the hurricane head for shore? If so, where will the hurricane strike? How bad will it be?

This method of modeling or computing the behavior of physical systems is called "finite-element modeling" because the size of each piece or element of the model mesh is finite (limited) in size. Finite-element modeling was developed in the 1950s for aircraft design and is used today in a many real-life applications, including machine design, ocean current modeling, medical imaging, climate change prediction, and others.

Real weather is not really made up of separate elements or pieces, but just as a picture made of dots can get more realistic as the dots are made smaller and more

numerous, a finite-element model can get more realistic as its elements are made smaller and more numerous. However, there is always a practical limit, because even the most powerful computers can only handle a limited number of calculations. In addition, some results, like hurricane forecasts, are needed in hours, not days or months. For this reason—and also because complex systems such as the weather are often "chaotic" or unpredictable over the long term by their very nature—weather forecasting will always remain imperfect. For example, when Hurricane Ivan approached the Gulf Coast of the United States in September 2004, computer models were unable to predict exactly where it would come ashore or how severe it would be by the time it did so. Repeated calculations starting from slightly different assumptions produced significantly different results.

SYNTHS AND DRUMS

Many of the characteristic sounds of modern pop music depend on math. From the deep "thump" of a driving hip-hop beat to the high-pitched "ting" or clashing sounds of high-hat cymbals, and lots of the sounds in between—snare drums, piano, strings, horns, hand claps, or sounds never heard before: are all produced mathematically, using functions.

Sounds are rapidly repeating pressure changes in air. If the pressure changes that we hear as a pure musical note are graphed, they look like a sinusoid (a type of function shaped like a wavy line). But music made only of pure sinusoids would sound flat and dull. More interesting sounds—sounds that growl, snap, clap, sing, and make our feet want to move—can be made by adding many functions together.

One of the ideas most commonly used in the design of sounds and sound systems is the fact, first discovered by French mathematician Jean Baptiste Fourier (1768–1830), that every possible sound can be built up by adding together the functions called sinusoids. Electronic sound synthesizers ("synths") allow a musician or sound designer to build sounds on this principle. In analog synths, an "oscillator bank" produces a number of sinusoids that can be shifted, amplified, and added to make new sounds.

This is not the only way that sounds can be created. Another method is called "physical modeling synthesis." In this approach, mathematical functions that describe the air vibrations made by a drum or other instrument are based the physical laws that describe the vibration of a drum head, string, soundboard, or other object. These functions are then evaluated by a computer to produce the desired sound.

NUCLEAR WASTE

Radioactive materials are made of atoms that break apart at random or unpredictable times. When an atom breaks apart ("decays"), it releases fast-moving particles and high-energy light rays. These particles and rays ("radiation") can kill or damage living cells. Only when most of the atoms in a lump of radioactive material have decayed is it no longer dangerous.

Radioactive waste is produced by the manufacture of nuclear weapons and by nuclear power plants (which make electricity) and are stored above the surface of the ground at about 130 locations around the United States. This is not a good long-term solution because accidents or terrorism might allow them to mix with the air and water. A possible solution is to bury the wastes deep underground in a part of the country that gets little rain, so that they can remain isolated until most of their radioactive atoms have decayed.

The time needed for radioactive material to become harmless is given by a mathematical function called an exponential function. This function states that after a certain amount of time—different for each radioactive substance—about half the atoms in a lump of that substance will have decayed. This amount of time is called a "half-life" of that substance. After another half-life has gone by, half of the atoms left after the first half-life will also have decayed, and so on, half-life after half-life, until the very last atom is gone. The amount of radiation given off by a quantity of radioactive waste thus decreases over time. The half-life of some radioactive substances is a fraction of a second; for others it is tens of thousands of years, or even millions of years.

If the amount of radioactivity at the beginning is R, then the amount after one half-life will be $1/2 \times R$. After another half-life it will be half of that, or $1/2 \times (1/2 \times R)$. If N is the number of half-lives that have passed, then the amount of radioactivity left will be R times 1/2 multiplied by itself N times, which can be written as $(1/2)^N R$. As a general rule, physicists say that a sample of radioactive substance can be considered be safe after 10 half-lives. By that time, the radiation will be down to $(1/2)^{10} R$, about 1/1,000 as much as there was at the start.

In the case of plutonium, a radioactive element found in nuclear waste, the half-life is about 20,000 years, so by this rule of thumb plutonium should be isolated from the environment for at least 200,000 years (10 half-lives).

These numbers are involved in a political dispute. Since 1978, the United States Department of Energy has been studying Yucca Mountain, in Nevada, as a place to bury about 60% of the nuclear waste that has accumulated around the country. Seventy-three miles (117 km)

Key Terms

Domain: The domain of a relation is the set that contains all the first elements, x, from the ordered pairs (x,y) that make up the relation. In mathematics, a relation is defined as a set of ordered pairs (x,y) for which each y depends on x in a predetermined way. If x represents an element from the set X, and y represents an element from the set Y, the Cartesian product of X and Y is the set of all possible ordered pairs (x,y) that can be formed.

Function: A mathematical relationship between two sets of real numbers. These sets of numbers are related to each other by a rule which assigns each value from one set to exactly one value in the other set. The standard notation for a function y = f(x), developed in the eighteenth century, is read "y equals f of x." Other representations of functions include graphs and tables. Functions are classified by the types of rules which govern their relationships.

Prime number: Any number greater than 1 that can only be divided by 1 and itself.

of huge tunnels to put the waste in have been dug 1,000 feet (305 m) below the surface at a cost of over $9 billion. However, no waste has yet been stored there. The site is still being studied, and government has decided that Yucca Mountain must be able to keep all its waste from leaking out for 10,000 years. Opponents of the plan—including the State of Nevada, which objects strongly to receiving the whole country's nuclear waste—argue that this is not long enough. In July 2004, a Federal Court of Appeals decided against the Department of Energy (which manages Yucca Mountain) in a lawsuit brought by the State of Nevada and environmental groups. The court ruled that 10,000 years was not long enough, and that the Department of Energy must come up with a tougher standard.

BODY MASS INDEX

Government experts said in 2000 that 11% of the United States population between the ages of 12 and 19 was overweight. That's more than double the rate measured in 1984; experts are talking of an "epidemic of obesity." But what is it, exactly, to be "overweight"?

Doctors decide whether a child or teen is overweight using something called the "body mass index (BMI) for children," also known as "BMI-for-age." Your BMI is your weight in kilograms divided by the square of your height in meters. If you are 1.7 meters (5 feet, 7 inches) tall and weigh 59 kilograms (130 pounds), then your BMI is

$$\frac{59\text{ kg}}{(1.7\text{ m})\times(1.7\text{ m})} = 20.4\text{ kg/m}^2$$

If you are heavy for your height, you will have a high BMI.

So is 20.4 a good BMI or a bad BMI? It's not necessarily either. Based on measurements of many thousands of young people, the Centers for Disease Control of the United States government have graphed average BMI as a function of age. That is, age is graphed as the independent variable (horizontal or *x* axis) and BMI as the dependent variable (vertical or *y* axis) to produce a curve. BMI is a good example of a function that is described not by an equation, but by a collection of number pairs. (There is one chart for boys and another for girls.)

When you go for a checkup, the doctor may compare your BMI to a chart to see how many people your age have BMIs less than yours. If 95% of people your age have a lower BMI than yours, you are considered overweight. If only 5% of people your age have a lower BMI than yours, you are considered underweight.

Where to Learn More

Books

Benice, Daniel D. *Finite Mathematics with Algebra*. Philadelphia, PA: W. B. Saunders Co., 1975.

Wheeler, Ruric E., and W.D. Peeples, Jr. *Finite Mathematics: An Introduction to Mathematical Models*. Monterey, CA: Brooks/Cole Publishing Co., 1974.

Web sites

National Centers for Disease Control. "BMI for Children and Teens." April 8, 2003. <http://www.cdc.gov/nccdphp/dnpa/bmi/bmi-for-age.htm> (September 16, 2004).

Game Math

Overview

Mathematics has played a role in games for centuries. Some games are purely mathematical in form, such as numbers theory, where solving a tricky math problem in of itself becomes the game. Other games, such as logic problems and puzzles, rely heavily on math to reach the solution. There are games where knowing a bit of math, such as probability, can help you determine your playing strategy, and still others where math helps you keep score.

Fundamental Mathematical Concepts and Terms

Game math covers a wide variety of entertainments, including board games such as Monopoly or chess, card games such as blackjack or poker, casino games such as roulette, logic puzzles, and number games. Number theory, a very specific sort of game math, deals with the make up of numbers themselves, and involves puzzling through the relationship between different numbers in order to find patterns in the ways that they relate. Often, number theory requires complex equations using algebra or calculus to come up with solutions, and in many cases there are mathematical questions that have yet to be answered—puzzles still unsolved. Games that involve the rolling of dice or the drawing of cards use addition, probability, and odds to determine a player's next move. Any game that involves keeping score requires someone to add up the points, and games that involve money require basic bookkeeping on the parts of the players. Computer game designers program their games to take into account the probability of each player's actions and the various reactions that the game might offer, working through every possible permutation in order to provide a realistic experience.

Probability and odds are two terms often used in relation to games that involve math, particularly games of chance. However, the terms are not interchangeable. Probability involves the outcome of a trial of chance in relation to the number of different outcomes that were possible. For instance, if you were to flip a coin, there would be two potential outcomes: heads or tails. Flip the coin once, and you will get one of those two choices. Therefore, if you flip a coin a single time, the probability that it will land heads up is one in two or 1/2. Likewise, the probability that it will land tails up is also one in two.

In the event that you have more potential outcomes, such as when you roll a single die, the principle remains the same. If you roll a die with six sides, numbered consecutively, there are six possible outcomes. The probability that you will roll a two is one out of six, or 1/6.

Odds refer to the chance that you will achieve a specific outcome compared to the chance that you will achieve any of the other potential outcomes. For instance, when tossing a coin a single time, you have only two potential outcomes, so the odds are that you will either have the coin come out heads up or tails up, one to one. These are also knows as even odds. But when rolling a single die, you have six different sides that might turn up. Therefore there is one chance of rolling any given number, such as two, while there are five chances that you will not roll a two, but one of the other potential outcomes. In this instance, the odds are one to five that you will roll a two, or 1/5.

In order to determine the odds against something happening, you take the mathematical reciprocal of the odds that it will happen. For example, in the instance of rolling the single die, if there is a 1/5 chance that you will roll a two, the odds are five to one or 5/1 against rolling a two in a single throw.

Permutations are the various choices at each stage of certain games, such as checkers, chess, or many computer games. Each time a player moves, they make a decision, and these decisions add up eventually to become their route through the game on that specific occasion. However, what would have happened if the player made a different move the first time? Or the second? In some instances there may have only been one choice, because an opponent blocked other routes or other moves led in the wrong direction, but ultimately, there were still a variety of options, and all of those put together are the potential permutations of the game. Mathematically, there are hundreds—sometimes thousands—of potential play scenarios available to each player, depending on how they mix and match their moves.

A Brief History of Discovery and Development

Game math has a long history, with some forms dating back to early human history. The throwing of dice originated in an ancient game that involved rolling bones, potentially man's earliest game of chance. Although it is unlikely that early gamblers were aware of the probability behind the game, it survived and eventually evolved into its more modern equivalent. Nor were dice games limited to specific nations. During his travels to China, Marco Polo reportedly encountered the dice rolling as both a means of divination and simple entertainment. Native Americans, the Aztecs and Mayans, Africans, and Eskimos all have evidence of dice in their cultures.

Early dice were fashioned out of animal teeth and bones, stones, and sticks, and used by witches or a tribal shaman to foretell the future. As they evolved and began to be used for diversion, the dice were shaped to match their uses. The modern look is believed to have originated in Korea, as part of a Buddhist game called Promotion. Dice as a means of gaming spread rapidly, particularly through the Roman Empire, where there is evidence not only of the rolling of dice, but of cheating. By the tenth century, dice games appear to have been a part of most cultures.

Chess is another mathematical game that has developed through the centuries. The earliest game resembling chess is Shaturanga, a board game between four armies that was designed by an Indian philosopher in the sixth century. The original version was played on a board made up of sixty four squares, and each of the four players had an army controlled by a rajah or king. Other pieces included infantry or pawns, calvary or knights, and an elephant that moved like the modern rook. In the early history of the game, dice were rolled to determine each player's moves. Only later, when Hindu law forbade gambling, were the dice eliminated. At that point, the game became a two-player contest, and the pieces were combined, with two of the kings demoted to prime ministers among other changes. The new version of the game was called Shatranj, and its first mention is in approximately A.D. 600 in a Persian text. As of A.D. 650, there was evidence of Shatranj being played in the Arab kingdoms, the Byzantine court, Greece, Mecca, and Medina.

Shatranj made its way to Europe some time during the eighth century, but there are several theories as to how. One theory is that the Saracens brought the game to Andalusia after conquering North Africa and it then traveled on to the court of Charlemagne in approximately A.D. 760. Another theory has Charlemagne engaged to the Empress Irene of the Byzantine Court. A Shatranj set was given to Charlemagne during one of their meetings, but instead of the standard prime ministers there were two queens mistakenly included. They were considered the most powerful pieces on the board and Charlemagne took it as a bad omen and cancelled the engagement. The third, and most likely explanation is that knights returning from the Crusades brought the game back to Europe with them.

The game continued to evolve across Asia and Europe, eventually breaking out into several versions. The European game most closely resembling modern chess became popular at the end of the fifteenth century. Certain pieces gained more power at that point, and several new moves were added to the game.

The common element through all of these versions of the game were the multiple permutations possible

depending on the strategy of the players. This became increasingly evident in modern times, when computer programs were written to simulate a chess game in the 1960s. The early programs were easy to beat, but as the programs became more complex, taking into account the possible mathematical permutations for each move and the likely responses of the opponent, the games became more sophisticated and more closely resembled the playing style of a live player. In 1997, Gary Kasparov, considered the best chess player in the world, was beaten by a computer chess program.

Magic squares are blocks of cells—three by three, four by four, etc.,—where each cell contains an integer and the integers in each row, column, and diagonal add up to the same number. Coming up with working combinations of numbers is a pastime that dates back to as early as 2200 B.C. The first known square of this sort was a third-order magic square, three cells across and three cells down, recorded in a Chinese manuscript. Early squares have been discovered engraved into metal or stone in both China and India. There is a legend surrounding the first magic square that says that Chinese Emperor Yu discovered it while walking by the Lo River. He spotted a turtle on the river's bank, and that turtle had a series of dots on its shell in the formation of a magic square, with each row, column, or diagonal of the dots adding up to fifteen. The Emperor took the turtle home to his palace to study it, and the turtle continued to live at court, with famous mathematicians traveling to examine it. The pattern of dots on the turtle's back became known as the Lo-shu.

Nearly every civilization since has made a study of magic squares, often attributing them with mystical properties and using them in rituals and as the foundation of prophecies and horoscopes. They appeared in Europe during the first millennium A.D., with the Greek writer Emanuel Moschopoulus being the first to write about them. The numbers have been equated with planets, elements, and religious symbols, and have appeared in works of art and on the backs of coins. As an entertainment, they have provided generations of mathematicians and laymen alike with a diverting puzzle, as people continue to attempt to discover new combinations of numbers that work in the magic square format.

Card games of all types also use a certain amount of math. It is unknown precisely when the first card games appeared, but there are references to card games in Europe starting in the thirteenth century, and the first actual playing card can be traced back to Chinese Turkestan in the eleventh century. It is suggested that the Chinese might have invented the cards, as they were the first to create paper.

Chess is essentially a game of mathematical options. WILLIAM WHITEHURST/CORBIS.

Specific card games, such as blackjack, are not referenced until more recently. Blackjack originated in France as twenty-one, or vingt et un, in the early eighteenth century. The game traveled to the United States in the nineteenth century and became popular in the West, where gambling was a growing pastime. Las Vegas legalized gambling in 1931, and blackjack became a regular attraction in the new casinos. By the 1950s, books were being written on how to count cards and predict the odds for any given hand of cards. Blackjack games began using more decks of cards, making it nearly impossible for an individual to keep track of the permutations, and in the 1970s, with the advent of mini-computers and calculators, cheating at cards rose to a new level.

Slot machines, which combine probability, statistics, and a variety of potential permutations and corresponding payouts, first appeared during the early part of the twentieth century. They were invented in 1895 by Charles Rey but not manufactured until approximately 1907 when Rey and the Mills Novelty Company joined forces to create the first machine, the Liberty Bell, which had a cast iron case and reels with pictures of playing cards. Different themes became popular through the years and other pictures were added, such as fruit, castles, and eagles. The modern

devices are electronic and provide even more potential for different combinations. Most illustrate the odds of achieving any given payout based upon the number of coins played. An understanding of probability helps a player realize the rate at which he will probably lose his money, and how slim his chances of hitting one of the jackpots.

Real-life Applications

CARD GAMES

Math is an important skill for various card games, both in order to strategize against one's opponents and as an actual part of certain games as well. In blackjack, basic addition is required in order to determine how to play each hand, in addition to an understanding of the value of each card in the deck. Numeric cards two through nine are worth the same number of points as their face value, so that a four is worth four points and so on. This is true of all four suits. The ten, jack, queen, and king are all worth ten points apiece. The value of the ace depends on the combination of cards in the hand, as it adapts itself accordingly, worth either one point or eleven, depending on which is more advantageous to the player. In order to get blackjack, the player's had must add up to exactly twenty-one points. This can happen as a natural blackjack—consisting of an ace and either a face card or a ten—or through another combination of cards.

There is more to blackjack than scoring a perfect twenty one, however. The game is usually played at a table with a dealer and one or more players. It is the dealer's duty to distribute cards both to the players and to himself—the house. In order to win, a player needs to come as close to twenty one as possible without going over, known as going bust, and yet still have either the same number of points or more points than the house. As a result, the game requires more math skills than simply adding up the cards in one's hand. Once the dealer distributes two cards to each player, the player must decide whether the cards he has are sufficiently close to twenty-one, or whether he wants to risk taking an additional card. Because the dealer only reveals one of his own cards at the onset of the round, the player also has to guess what the chances are that the dealer has a better hand. The player's only advantage is knowing that the dealer must continue to take additional cards until the house's hand reaches at least a total of sixteen. The house must also stand, or take no additional cards, once the dealer has reached a total of seventeen points in the hand.

In order to play blackjack effectively, a player must determine the mathematical likelihood that they will have a hand equal to or better than the dealer's hand, without going bust. There are some basic rules of probability that come into play at this point. For instance, looking strictly at the player's hand, if the two cards dealt total seventeen or higher, it is advisable for the player to stand and refuse any further cards. It is easy to see why when you look at the possible permutations. If a player already has a total of seventeen and chooses to draw an additional card, the only cards that would give him a new total of twenty one and lower are the ace—acting as a one and not eleven—or the two through the four. Five or higher will result in the player going bust. This means that out of a potential thirteen cards in a suit, four of them would result in a win, while the remaining nine would result in a loss. Not very good odds.

The math becomes more complicated when the player's original hand totals less than seventeen, and it becomes necessary to also consider the value of the single card the dealer has revealed for the house. The higher the dealer's standing card, the more likely it is that when the second card is revealed the total will be above seventeen. Therefore, the higher the dealer card, the more likely it is that the player will need to choose an additional card for his own hand. This becomes doubly true if the two original cards dealt to the player are substantially lower than sixteen. Once the various permutations are taken into account, it becomes clear that if the player's hand totals twelve or less and the dealer holds anything other than a four through a sex, the odds are better if the player draws an additional card. If the player holds thirteen to sixteen points, and the dealer holds a seven or higher, the player should also draw another card. It is important to remember when totaling a hand that an ace can count as either a ten or a one, so an ace and a six can be considered either a sixteen point hand or a seven point hand. Also, the probability of drawing any given hand changes depending on how many decks the dealer is using and how many hands have already been played.

Counting cards is a skill that can help a player get a better idea of his chances of winning over the course of a game. The system involves assigning each card a value other than their point value in the game. The ace and any card with a point value of ten is equal to negative one, cards two through six are equal to positive one, and the others are counted as zero. For each hand, the player keeps a running tally based on the cards dealt and the number of decks being used. There is a higher chance of winning when the card count total is above a positive two.

In other card games, counting cards merely refers to keeping track of what has already been played. If a single deck of fifty two cards is divided among four players, each

player guesses what the other players are holding based on their knowledge of their own hand and their knowledge of the cards in the deck. Depending on the game, as cards are thrown down on the table, a player remembers what they have seen because that means those particular cards have been eliminated from play. The player, in a sense, subtracts those cards from the total set available.

Probability is also used to determine what sort of hand constitutes a winning hand in many card games. For instance, in Poker, a hand that contains a pair of like cards may have some value, but it will still be worth less than three of a kind or a flush, and far less than a royal flush. The probability of getting a pair is fairly high, given that there are more than one million potential combinations in a standard deck of cards that would qualify. A royal flush, however, which requires you to not only attain the major face cards but for all of them to be in the same suit, is limited to only four potential combinations—the face cards for hearts, diamonds, spades, and clubs. It is easy to see why a royal flush is such a difficult hand to obtain, and why it will beat out any of the other potential combinations.

OTHER CASINO GAMES

Game math in the form of probability plays a huge role in most other casino games as well. Roulette is an obvious example. A player chooses a number on the board and places their chips on that square. Once all of the bets are made, the dealer spins the wheel, setting loose a small ball that circles the wheel until the wheel comes to a halt, then drops down into one of the slots that indicate the winning number. It is similar to playing the lottery, but on a smaller scale. In order to win, a player must have bet on the same number as the one on which the ball lands.

Other bets are possible in roulette, but as with anything, the greater the odds of winning, the less money you lose—or win. Half of the numbers in Roulette are black while the remaining half are red. If a player decides to simply bet on black in general, they have a fifty/fifty chance of winning and therefore can do no more than double their bet with each spin. At the other end of the spectrum is a straight up bet, where a player chooses a number and hopes that is the one to win. Because there are thirty six numbers on the board, the odds of winning become one in thirty five or 1/35. In between these two bets are other combinations that provide various different odds, such as a trio bet in which three individual bets are placed on separate numbers and the odds of winning become three in 36 or one in 12.

It is important to understand the odds of a game before playing because the mathematics behind the action can help a player from making foolish mistakes. It is possible to place a bet in Roulette thirty six times in a row and never win, even though there are thirty six numbers, because the win spins independently each time and each time the odds of a given number appearing remain virtually unchanged. Common sense tells us that eventually different numbers will result, but experience teaches us that it is unlikely that each of the thirty six numbers will come up once. However, if a player takes those same thirty six bets and places them all at once, with one bet on each number, the outlay of chips is identical but there is a guarantee of winning because all of the potential outcomes have been covered.

Slot machines offer players the chance to play with minimal interaction. All one needs to do is put money into the machine and either pull the lever or push a button. If the machine offers the opportunity to bet more than one coin at a time, the player can also make that decision. However, there is no other human interaction—the rest of the game is up to the machine. Players win at various levels based on whether any matched sets appear on the reels when they stop spinning, but these wheels have been programmed and are encased within the machine, so a player has to trust to the payout odds listed on the front of the device to determine whether the probability of winning is worth the risk of losing. As with many casino games, the odds are in favor of the casino.

The payout percentage is listed on most slot machines. This is the amount of money that particular machine is required to pay back out in winnings over the course of its lifetime on the casino floor. However, a high percentage, while favorable, is no guarantee, as the machine could pay that entire amount in one huge jackpot every so often and then not provide any smaller winnings in the interim, or conversely, could pay the money in small, unimpressive amounts on a steady basis. There is no way to know the machine's history, or whether it might have paid out a fairly large amount of money to another player just an hour earlier. The machine should, however, illustrate some of the typical combinations that might appear when the reels stop spinning, and what the odds are of that type of payout occurring. It will also show if the machine accepts multiple coins for each spin, and what the difference is in the payout if you play one, two, or more coins.

Multiple line payout machines seem to offer multiple chances to win by paying for lines other than the one straight across the middle. For instance, a three line machine might offer payout for a line across the top, middle, and bottom of the reels. However, often these lines only payout if you are making the corresponding bet, so

Nigel Downing poses with his board game, "Enterprise Profit Ability." The board game teaches players how cash flow and profit affect business. AP/WIDE WORLD PHOTOS. REPRODUCED BY PERMISSION.

that one coin activates one line, two coins activate two lines, and so on. A player has to take a close look at the probability of their winning and determine if it is worth playing several coins at a time.

Progressive machines are a relatively new type of slot machine. A progressive machine reserves a small amount of money from each bet made and adds it to a grand jackpot total that continues to grow until someone finally wins. What makes this impressive is that multiple machines are linked together, with all of them feeding into the one prize total. They can be part of a local network, with machines all over that particular casino or even several nearby casinos under the same management, or they can be part of a wide area network, which could include slot machines over a fairly large region. This means that the jackpot has the potential to be a far larger amount than if it were just based on a single machine. However, because money is constantly reserved to add to that total, the odds of winning are greatly reduced, and any non-jackpot winnings tend to be extremely small. Also, many progressive machines require players to bet the maximum number of coins in order to win the progressive, so if the jackpot comes up and a player has only bet a single coin out of a possible three coins, there won't be a payout.

BASIC BOARD GAMES

Board games rely on several mathematical principles to enable the players to advance around or across the board. Simple games for children, such as Chutes and Ladders and Candyland, provide the players with an arrow to spin, and whatever number the arrow lands on is the number of spaces the player can advance. This basic system enables younger children to grow accustomed to counting and adding as they work their way through the game.

More advanced games for older children and adults often use dice to determine the number of spaces a player advances. Again, they use simple addition, but often take other factors into consideration as well. It is common for a game to require players to roll a specific number at some point during the game, whether it is double sixes for bonus points or an exact number that allows the player to reach the final square of the game in the precise number of moves.

Monopoly provides players with numerous chances to apply their math skills. Not only does this game use dice to travel around the board, but it doles out money to each player based on various events over the course of the game and also requires players to pay out from their accumulated earnings in order to advance. For instance, players receive a two hundred dollar salary each time they pass the initial starting point on the board, the "Go" square. They can also land on squares that allow them to draw cards that occasionally reward other cash prizes, such as a lottery win or an inheritance. Expenses for the player fall into two categories: those that are unavoidable and those that serve as an investment. If a player lands on a square—or property—owned by one of his opponents, he must pay rent to the owner of that property. Players can also lose funds by landing on squares that require them to pay taxes or alimony or to go to jail and forgo passing "Go" and collecting that round's salary. However, if a player lands on a property that is not owned by another player, he has the option of buying it himself. This costs money from his own fund, but provides him with the potential to earn additional cash if and when his opponents land on that space. Rents are further escalated when players manage to buy several properties of a set, and when houses or hotels are purchased to add to the value of the square.

Chess Mathematics

Chess is a highly mathematical game, as the only way in which to excel is to study the board and determine the various potential moves for each piece in play—both your own and those of your opponent—for several moves into the future. This understanding of probability and permutations allows a player to strategize for the best outcome. The chess moves themselves, when written out, use a form of algebraic notation to explain what piece has been moved to which square on the board, and the board is referenced as a grid.

However, there is more to chess than the traditional game of white against black. Within the game itself is another puzzle called the Knight's Tour.

The Knight's Tour involves traveling a knight using only the traditional combination move of two squares either horizontally or vertically and one square at a right angle to that move so that the knight lands once and only once on each square of the board. This puzzle has likely provided an intriguing challenge nearly as long as Chess itself has existed, and is referenced as early as one thousand years ago. The first in-depth study of the math behind the puzzle was published by the mathematician Euler in 1759.

There is no rule that a Knight's Tour must take up the entire traditional board comprising sixty-four squares. Smaller puzzles exist, and in some cases the smaller blocks are solved and then combined in an effort to find a new way of traversing the larger board.

Math puzzles can be combined, so that a Knight's Tour forms a quasi-magic square. In order to do this, each move the knight makes is numbered consecutively, with the number written into the landing square. Once the tour is complete, the numbers are then added across and down to make a partial magic square.

As difficult as the Knight's Tour is, there is an even more challenging version, known as a closed Knight's Tour, in which the knight finishes the tour of the board on the same square from which it began. This is also referred to as re-entrant. Other ways to make the tour more difficult include turning the flat chess board into a cube, where each face is a separate board of sixty-four squares and the knight must tour each face in such a way as to ensure that the last move allows it to jump to the next side of the cube.

Because the knight has the most complex way of moving around a chess board, it is the subject of a variety of these types of mathematical puzzles. Another poses the question of how many knights would need to rest on a board in order that every single square would be reached in a single move. On a standard chess board, the fewest number of knights required would be twelve.

MAGIC SQUARES

Magic squares are a form of math puzzle that dates back thousands of years. A magic square is made up of a number of cells, an equal number per row as per column, with an integer in each cell. The integers used are the numbers 1, 2, 3, . . . N^2, where N is the number of rows or columns. When added together, the sum of each row is equal to the sum of each individual column, as well as the sum of each of the diagonals. Magic squares are essentially puzzles with very little in the way of practical value. However, the patterns are often used in art and geometric designs.

The smallest possible magic square consists of nine cells—three across by three down. A magic square of two cells by two cells would only work if every cell held the number one, and so is considered too simplistic and therefore not a true magic square. The number of cells in a row or column, N, determines the order of the magic square. Therefore, a square consisting of three cells by three cells is a third order square. The order of the square also determines what the sum of each row and column will be. Every third order magic square consists of rows, columns, and diagonals that add up to the number fifteen. A fourth order magic square—four cells by four cells—consists of rows, columns, and diagonals that total thirty four.

In order to determine what the sum is for a particular order of magic squares, all you have to do is plug the order number into a simple equation. For instance, assume you are interested in knowing the sum of each line for a fifth order magic square. N = the order of the square, so in this case N = 5.

$S = (N/2)(N^2 + 1)$ where S is the sum for any row or column in the magic square. In this example, for a fifth order magic square, the sum is sixty five. The sum is often referred to as the magic sum or magic constant.

But how do you determine what numbers go on which lines so that each of the lines adds up properly?

With small orders of magic squares, it is easy to see how to distribute the numbers because there are only a few potential equations to use. For example, in a third order magic square, where each line totals fifteen and each cell holds a single digit number, one simply examines the various different equations that meet those terms. Each equation needs to consist of three numbers. So, for a third order magic square, here are the possible equations:

$6 + 5 + 4 = 15, 7 + 5 + 3 = 15, 7 + 6 + 2 = 15, 8 + 4 + 3 = 15,$

$8 + 5 + 2 = 15, 8 + 6 + 1 = 15, 9 + 4 + 2 = 15, 9 + 5 + 1 = 15$

In this instance, the first thing that becomes apparent is that the number five appears in four of the eight potential equations for the third order magic square. This means that the number five should be placed in the center cell of the square, since four separate equations are required to use the center cell as one of their integers: the middle column, the middle row, and each of the two diagonals. Once those equations are in place, it is simply a matter of determining how the other cells need to be filled to match the equations as listed above.

The number of potential magic squares increases with each increase in the order. There is only one true third order magic square, although it is possible for it to appear as if there are more, depending on which way one flips the square itself. However, there are 880 fourth order magic squares, and over thirteen million fifth order magic squares. Obviously, one cannot simply list potential equations in order to determine how to fill the cells when the possibilities become so numerous. However, there is a way to generate a new magic square based on the layout of an existing one. For each cell in the magic square, subtract the integer from N2 + 1, then insert the new number into the cell. The new magic square is sometimes called the complement of the original square.

There are different types of magic squares. Some refer to the classic magic square described above, where the numbers in the cells consist of the integers from 1 through N^2, as a normal magic square. A magic square that allows other integers in the cells as long as the sums of the rows, columns, and diagonals work is then referred to as simply a magic square. Creating normal magic squares, however, is by far the more challenging puzzle.

In a normal magic square in which N is equal to an odd number, there are several systems for filling in the rest of the integers. Perhaps the simplest is called de la Loubere's algorithm. This method has you start by placing the number one in the cell that is at the top center of the magic square. Then working in numerical order, you add each number to the square by working in an upward diagonal direction to the right. So, after the number one, you move one cell up and one cell to the right and place the number two in that cell. Of course, because you have commenced the square at the top, this movement takes you out of the square entirely. In order to determine where the number two should go, you must imagine that the magic square repeats in all directions. Once you've filled the cell in with the number two, and have determined where in the square it goes, you can transfer it to the same position in the original square. In the case of a third order magic square, the two would therefore appear in the bottom right hand cell.

De la Loubere's algorithm enables you to continue filling in the cells of the magic square until all of them have been filled. In some cases, the system of moving one cell up and one cell to the right will lead to a cell that has already been filled. When this happens, you simply drop straight down by one cell from the most recently filled cell and insert the next integer there instead. Then resume the standard movement of up one and over one to the right with the next consecutive number. In the third order magic square, this leads to the number four appearing below the three, and the seven beneath the six. This system also works if you spread the move out to resemble a knight's move in chess, in which you move the piece two cells either horizontally or vertically and then one cell at a right angle to the original direction. For the purpose of the magic square, the movement must always consist of two cells up and one cell to the right.

There is no algorithm to create magic squares where N is equal to an even integer, however. Despite this, many have been discovered. Perhaps the most famous even integer magic square is the fourth order square that appears in artist Albrecht Dürer's etching, "Melencolia I." The German Renaissance artist drew the magic square into the top right hand column of the work. Astrologers of that period linked fourth order magic squares to the planet Jupiter, and they were thought to battle melancholy. As the etching depicts a woman thinking while surrounded by uncompleted chores, it is possible that this was Dürer's purpose in including the magic square.

The square itself, however, has certain intriguing properties. The cells read across as 16, 3, 2, 13 for the first row, 5, 10, 11, 8 in the second, 9, 6, 7, 12, in the third row, and 4, 15, 14, 1 along the bottom row. As expected with a fourth order magic square, each row, column, and diagonal adds up to thirty four. In addition, however, the four corner numbers (16, 13, 4, 1) also add up to thirty four, as do the four numbers that comprise the inner square of two cells by two cells (10, 11, 6, 7). The sum of the remaining numbers is equal to sixty eight, which is twice the standard sum for any given line. As an added

detail, the two central bottom cells of the square read 15 and 14; 1514 was the year in which Dürer created the etching.

MATH PUZZLES

Math puzzles or logic puzzles frequently consist of a short story in which a problem is given and the aim is to solve the puzzle through math. In a school setting, they are often called word problems, but they have a history of being offered up as challenges to see who can find the conclusion first—or in some cases, at all. Often these problems combine math with common sense to see if the person attempting to solve it is paying attention to all of the details.

Here's an example: an injured mouse is trapped at the bottom of a hole that is ten feet deep. Each day, the mouse is able to climb up three feet, but each night he slides two feet back down. How many days will it take the mouse to get free of the hole?

Well, at first glance, it seems that his is a simple math problem. The hole is ten feet deep. Every twenty-four hours, the mouse travels three feet forward and two feet back, which means he makes precisely one foot of progress per day. Therefore, the automatic response would be that, at one foot per day, it would take ten days for the mouse to travel ten feet and thereby climb out of the hole.

Except, it's not quite that straightforward because the mouse does make that three feet of progress every day, even if it does then lose a good portion of it during the night. Day one sees the mouse crawl to the three-foot mark, then slide down, so he starts day two at the one foot mark. He then crawls to four feet and slides down to two. And so on. On day eight, however, the mouse is starting at the seven-foot mark, because that is where he slid to the previous night. But when the mouse crawls three feet up, he has reached the ten foot mark, thereby reaching his goal and climbing out of the hole.

Other traditional math puzzles such as this include problems where the ages of several people are given but only in relation to each other, and the aim is to determine how old each of them are; time and distance puzzles, where one must try and determine which train will reach the city first based on when each leaves in relation to each other and how far they are traveling at a particular speed; and puzzles where an unstated number people choose items from a box and based on which items and how many each takes, one must determine how many people there actually are. These games are enjoyable for sheer entertainment value, but also help develop problem-solving skills that can be used in the real world.

Potential Applications

Game math serves purely as entertainment if you look at it as a form of recreation, but it can also mean big business if you are on the other end of the spectrum. The casino industry relies heavily on a solid knowledge of probability and the ways in which they can keep the odds in their favor, as that is how casinos make their money. A casino needs to keep players entertained and interested, or they will not continue to gamble. Therefore, the players must be able to win often enough that they are having a good time, but not so frequently that the casino ceases to make money. The casino's goal is to earn as much of a profit as possible while still keeping their customers happy. In order to keep everything in balance, they analyze the odds on every game on their floor and adjust them whenever necessary. Mechanized games, in particular, can be manipulated to pay out at a different percentage rate to keep players entertained and coming back for more.

Probability comes into play in any game that enables you to place a bet, including horse racing and other professional sports. Odds are determined based on numerous factors including previous performance results and other analysis. As with any form of gambling, the higher the odds one of the participants might win, the greater the potential payout, and the greater the chance of losing one's money.

Computer games are another enormous industry that relies heavily on various types of game math. Interactive games that allow you to choose a character's moves based on a set scenario are programmed with thousands of different permutations to take each possible choice into account. Each individual choice leads down numerous different paths, and each one results in a different ending. The best game designers understand that a variety of options makes the game more entertaining and insures that the players will be able to play over and over instead of simply solving the puzzle once and having to move on to something else.

Where to Learn More

Books

Falkner, Edward. *Games Ancient and Oriental, and How to Play Them*. New York, NY: Dover Publications, Inc., 1961.

Hunter, J.A.H. and Joseph S. Madachy. *Mathematical Diversions*. Princeton, NJ: D. Van Nostrand Company, Inc., 1963

Pickover, Clifford A. *The Zen of Magic Squares, Circles, and Stars*. Princeton, NJ.: Princeton University Press, 2002.

Web sites

Casinos and Gambling History. <http://www.worlds-best-online-casinos.com/Articles/History.html> (May 6, 2005).

Key Terms

Odds: A shorthand method for expressing probabilities of particular events. The probability of one particular event occurring out of six possible events would be 1 in 6, also expressed as 1:6 or in fractional form as 1/6.

Permutations: All of the potential choices or outcomes available from any given point.

Probability: The likelihood that a particular event will occur within a specified period of time. A branch of mathematics used to predict future events.

Delphi for Fun. "The Knight's Tour." <httphttp://www.delphiforfun.org/Programs/knight's_tour.htm> (May 7, 2005).

E-How. "How to Count Cards." <http://www.ehow.com/how_4369_count-cards.html> (May 6, 2005).

Interactive Mathematics Miscellany and Puzzles. <http://www.cut-the-knot.org/games.shtml> (May 7, 2005).

Magic Squares. "Vignette 20." <http://www.jcu.edu/math/vignettes/magicsquares.htm> (May 8, 2005).

Math Forum. "Magic Squares." <http://mathforum.org/alejandre/magic.square.html> (May 7, 2005).

Math World. "The Knight's Tour." < http://mathworld.wolfram.com/KnightsTour.html> (May 7, 2005).

Math World. "Magic Squares." <http://mathworld.wolfram.com/MagicSquare.html> (May 7, 2005).

Mathematica Gallery. "Distribution of the Knight." <http://www.tri.org.au/knightframe.html> (May 7, 2005).

Online Guide to Traditional Games. <http://www.tradgames.org.uk/games/Chess.htm> (May 7, 2005).

The Knight's Tour. <http://www.borderschess.org/KnightTour.htm> (May 7, 2005).

Wizard of Odds. <http://www.wizardofodds.com> (May 7, 2005).

Game Theory

Overview

Game theory is an approach to the way humans interact and behave that is rooted in mathematics. Game theory attempts to understand, explain, and even to predict the "give and take" between people and other organisms (even bacteria follow game theory) that allows an outcome to be reached.

Game theory does not involve entertaining games in the manner of tag or cards. The game in game theory, however, resembles tag or poker in that decisions have to be made to reach an outcome, and there are winners and losers based on the decisions that are made. Also, like tag and poker, game theory operates on the premise that the two or more people who are involved have an interest in the outcome. In most cases, the participants want to win. It is the strategies used in trying to achieve a winning outcome that game theory helps explore.

Some aspects of human behavior lead to individual gain. By performing in a certain manner, an athlete could attract the attention of the basketball coach, and so increase his chances of making the starting squad. One person making the team, however, comes at the expense of someone else not making the team. This area is part of non-cooperative game theory. The subject acts on his own, for his own benefit.

At other times, chances of success come when people get together and pool resources with others. This is called mutual gain. For example, in the basketball analogy, after a person makes the team, chances of success (winning games and being a star) happen only if that person works together with his teammates. By cooperating, the team wins and everyone benefits. The part of game theory that looks at this sort of behavior is known as cooperative game theory.

Game theory was first developed to help those who study the economy learn how decisions are made in the face of conflict. Whether in a small business deal or in economic relationships between two nations, conflict is one of the driving forces of the proceedings. By understanding how rational decisions that benefit the person or organization making them can be made in such an atmosphere, the chances of making a good decision improve.

As the power of game theory became more recognized over time, the theory was applied to other areas such as sociology, psychology, biology, and evolution. Game theory is also important in sports, mathematics, social science, and psychology (psychologists often refer to an aspect of game theory as the theory of social situations). It even appears that evolution operates according

to aspects of game theory. A genetic change is tried out in the face of a survival pressure. If there is a benefit to the change, it is conserved and the change continues.

Game Theory Issues

The following issues, posed as questions, illustrate the relevance of game theory. Some might seem familiar:

- A strategy is chosen to achieve a desirable outcome. How can this strategy be chosen rationally when the outcome depends on the strategies that are selected by others and when you do not have all the available information on the topic when you select your strategy?
- When a situation permits all those involved to gain (or to lose), is it rational to cooperate with the others to realize the mutual gain or mutual loss, or is it more rational to act independently and aggressively in order to claim a larger share, even at an increased risk of loss?
- If it sometimes beneficial to cooperate with others and sometimes beneficial to act independently, how are you to know which decision is the rational choice at a particular time?
- Do all of the questions above apply in the same manner to an ongoing relationship as compared to a one-time relationship?

These questions help to point how the important features of game theory. It is about how people get along with one another as much as it is about how we compete with each other. If we were completely competitive, such a predatory environment probably would have ended humans as a species long ago. We thrived because we learned to cooperate.

Game theory, therefore, goes to the heart of human nature, illustrating cooperation and independent behavior at different times. To understand which course is best at a particular time is what game theory is all about.

Fundamental Mathematical Concepts and Terms

The mathematics that govern game theory attempt to model and predict the outcome of interactions between people, or, in a field like evolutionary game theory, between an organism (such as a bacterial cell) and something else (such as an antibiotic). As summarized above, there are two main branches to this interaction: the cooperative game theory and the non-cooperative game theory.

The thing that drives game theory is the outcome. Most of us want the outcome of a decision to be something that is good for us, makes us feel happy, and which benefits us. In the jargon of the game theory world, this good stuff is called the "payoff."

In game theory, there are decision makers (the players) who each have at least two choices or a defined series of choices (the plays). All the different combinations of plays leads to some end (win, loss, or draw), and this ends the 'game.' Some games end when one player wins and the other loses at the same time. This is called a zero-sum game. Capturing your opponent's Queen in the game of chess is a perfect example of a zero-sum game.

Chess also provides another example of an outcome, namely a draw. Both players may realize that winning is not to be. So, they cooperate to end the game by deciding that there will not be a winner or a loser.

In game theory it is assumed that each decision maker has all the necessary and important information needed for the decision. Everybody knows as much as everybody else. As we saw above, this is not always the reality. However, it makes for a starting assumption. Furthermore, it is assumed that all players make their decisions rationally (for an interesting twist on this assumption, see "The Prisoner's Dilemma").

A Brief History of Discovery and Development

The popular root of game theory dates back to 1944. It was then that the renowned mathematician John von Neumann, in collaboration with the mathematical economist Oskar Morgenstern, published a book called *The Theory of Games and Economic Behavior*. This book laid out the framework of game theory upon which others have added to over the years.

However, the roots of game theory go back much further than the middle of the twentieth century. For example, the Babylonian Talmud gathered all the then-existing laws and traditions as a basis of Jewish religion, criminal law, and social interactions. One recommendation (Mishna) concerning the division of marital property upon the death of the husband among his wives (more than one being the norm at that time) has various options depending on the size of the estate. In the case, three wives whose marriage contracts specify that in the case of this death they receive proportions of his estate of 100, 200, and 300, respectively (there being some sort of seniority in place). If the estate was only 100 (a relative figure for the purposes of the example), the Talmud

The Prisoner's Dilemma

This classic example of game theory was first devised in the 1950s by two researchers of the RAND Corporation, when the "Cold War" threat of nuclear weapons was part of everyday life. The game was an attempt to understand decision making in times of stress.

The Prisoner's Dilemma relates the story of two prisoners (we will call them Fred and John). In the story, Fred and John team up to rob a store. Because they both planned the heist, they both know all the details. Later they are both picked up by the police under suspicion of being the men who committed the crime. They are carted downtown to the police station, and are put in holding cells that are some distance away from one another. There is no way they can communicate with each other.

A prosecutor offers Fred and then John a break on the length of their jail sentence if they give him information of the other man's participation in the holdup. He tells Fred that he has offered the deal to John, and tells John that he has offered the deal to Fred.

The deal offered to Fred and John is this:

- If one of the prisoners confesses that the two of them committed the crime and the other prisoner denies that he had anything to do with the heist, then the one who confessed will be set free and the one who denied any wrong doing will get a 5-year prison sentence.
- If both Fred and John deny doing anything wrong, they will probably be convicted and each receive a 2-year jail sentence.
- If both Fred and John confess that they committed the crime as a team, they both get a 4-year sentence.

The Prisoner's Dilemma now shifts to consider one prisoner. Let us consider the case of Fred. If Fred implicates John and he denies having anything to do with the convenience store holdup, he is free. However, if Fred says that John was involved and John tells of Fred's involvement, both get a jail sentence of four years. Finally, if Fred does not mention John's involvement but John says that Fred was a part of the robbery, Fred is off to jail for five years.

It is to Fred's advantage to implicate John, and hope that John tells the police that Fred was also a part of the crime, and so go to jail for a shorter time than if Fred denies involvement hoping that John also denies involvement. This is because Fred runs the risk of John implicating Fred while Fred denied doing anything, and thus Fred would be up for a 5-year prison stay.

The action		Jail sentence (the payoff)	
Fred	John	Fred	John
cooperate	cooperate	2 years (R)	2 years (R)
cooperate	deny	5 years (S)	0 years (T)
deny	deny	0 years (T)	5 years (S)
deny	cooperate	4 years (P)	4 years (P)

Table 1.

The rational thing to do is implicate John. Of course, John arrives at the same decision. They both end up confessing and so both go to the state penitentiary for four years. However, if both denied the crime, they both would be facing jail time of only two years.

By acting rationally, both have come out worse! It turns out that Fred and John's best strategy would have been to act irrationally and admit that they did do the crime. In this game, cooperation is both the best thing to do and an irrational thing to do!

The Prisoner's Dilemma explains how this has come about in terms of what is called a payoff (the consequence of the action). There are four payoff categories:

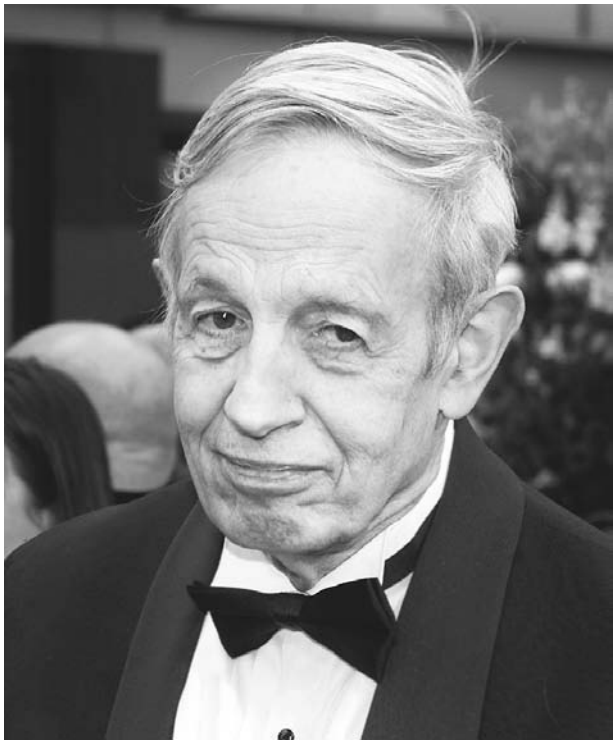
- R = reward for mutually cooperating
- S = sucker (admitting guilt thinking that the other person will do the same)
- T = temptation to deny
- P = punishment for mutual denial.

The dilemma is set up so that the rank of these preferences, from the most desirable to the least desirable, is T, R, P, and S. As well, the reward payoff R can be greater than the average of T and S payoffs.

Table 1 presents a chart that can help make things clearer to understand.

Another fascinating aspect of the Prisoner's Dilemma is the outcome if the same participants 'play' again. The next time Fred and John wind up in the jail together and get a offer of time off for spilling the beans on each other, Fred could well select a different option, based on your memory of John's response the first time around. So, the game becomes more complex as the same people play it over again. Likewise, if more than two people are involved, there are more interactions that can occur, which also increases the complexity.

The Prisoner's Dilemma model has been applied to many real-life situations, from the way business is conducted to political relations between countries, and even to how bacteria deal with the presence of an antibiotic.



John Forbes Nash PHOTOGRAPH BY ROBERT P. MATTHEWS. © REUTERS NEWMEDIA INC./CORBIS. REPRODUCED BY PERMISSION.

recommends that the estate be divided in three equal portions among the three wives. However, if the estate totaled 300, then the wives divide up the property in a different way, in a ratio of 50, 100, and 150 (the numbers in each example do not add up). Finally, while for an estate that totals 200, the Talmud recommends a proportional division of 50, 75, and 75. The reasons for the different divisions, and how the ratios were arrived at was, for a long time, a confusing mystery. But, in the 1980s scholars recognized that the options were based on what we now know as cooperative game theory. Each of the solutions was a logical response to the given situation (or 'game'). Thus, even in A.D. 500, game theory guided decisions.

Another milestone in games theory made it to the Hollywood screen. The movie *A Beautiful Mind* the Oscar-winning performance by Russell Crowe chronicled the troubled genius of Princeton University researcher John Forbes Nash. Between 1950 and 1953 Nash produced four papers that proved to be of major significance to game theory and made game theory very useful to non-cooperative situations like the bargaining process that goes on between workers and management in seeking a new work contract. Subsequently, Nash went through decades of torment, due to the development of schizophrenia. Ultimately, he was able to deal with his mental illness. The brilliance of Nash's insights culminated with his being awarded a Nobel Prize in economic sciences in 1994.

In the mid-1950s, game theory was applied to the political arena. Two researchers (L.S. Shapley and M. Shubik) developed and used a calculation (the Shapley value) to determine who in the United Nations Security Council wielded the power. At about the same time the link between game theory and philosophy was recognized.

The 1960s also saw game theory applied to automobile insurance, where the rates that are set are influenced by the degree of risk.

Just several decades later, in 1982, a book entitled *Evolution and the Theory of Games* was written by John Maynard Smith. In the book, Smith applied game theory to evolutionary biology; the inherited biological changes that are driven by evolutionary pressures.

The intervening years have seen the applications of game theory expanded still further, and the tools used to apply the theory become more refined.

Real-life Applications

ECONOMICS AND GAME THEORY

A major application of game theory is to economics; the generation of wealth, creation of jobs, and the flow of money and goods that keeps societies from collapsing. Economic game theory also has three other related areas to consider. These are known as decision theory, general equilibrium, and mechanism design theory.

DECISION THEORY

You are golfing. You are getting ready to hit your drive on a particularly hard hole. The reason that this hole is so challenging is a pond that cuts across the fairway about 250 yards (228.6 m) away. If you hit a good shot, your ball will clear the pond. However, your shot could just as easily get wet. You could have a go at bashing your ball over the pond and leaving yourself a really easy second shot to the green. But that requires a pretty good shot. A not-so-good shot will be at the bottom of the pond, and you will have an added penalty stroke. Maybe you should play it safe and hit a shorter shot that does not make it the pond. That leaves you with a longer second shot to the green. How do you feel about your golf skills today? What are you going to do?

This example, which involves one person thinking about the particular situation, acquiring information, and using the information to make a decision that

determines his or her outcome, is what decision theory is all about.

GENERAL EQUILIBRIUM

General equilibrium is not on the personal scale, like our previous golfing example. It is much larger in scope. The concept is suited to making or getting products to a large number of people. This can be literal, as in the manufacture of something and the distribution of the item to stores nationwide. However, general equilibrium can also be used to consider things like the stock market, and even politics.

NASH EQUILIBRIUM

The Nash equilibrium is named after John Nash. The premise that he developed (when he was still a graduate student, and before mental illness claimed him for several decades) is that participants in an activity have a number of options. The equilibrium exists when no one has any reason to change their selected option, since by doing so they earn less (Nash was thinking about economics) than if they hold their course. The outcome, however, does not have to be about money.

So, when someone contemplates a new strategy as a way of earning more or maximizing their payoff, he/she needs to consider what the others are doing and how they might change what they are doing. In a Nash equilibrium, the best individual response is to cooperate.

ECONOMICS

Economics and game theory go hand-in-hand. Indeed, game theory came about as a way of getting a handle on economic activities.

The link between economics and game theory is rational behavior. Although economists can disagree in the nuts and bolts of economic theory and which economic plan is best, the general consensus (often referred to as neoclassical economics) is that people are rational in their economic choices. We consider our options and make the choice that we feel gives the best opportunity for the best result. When the aim is to get maximum return on an investment, or to make the most money we can make out of an opportunity, a rational approach is certainly the wise approach.

If we were operating in a vacuum, with no outside forces affecting our economic decisions, then the decisions would be easier to make. However, life does not operate so simply. A person's decision is influenced by, and the outcome of a decision affected by, factors like political relationships between countries, the stock market, company fortunes, and the changing currency rates in your country and around the globe.

Game theory arose to enable economists to predict the outcome of an individual's decision in the face of all these unpredictable influences. In game theory, as in a real-life game like poker, a person chooses a strategy. The outcome of this choice depends on the strategies that are chosen by other participants.

Game theory applies to economics in more ways than just making money. The process of pondering a purchase can be guided by game theory. For example, a student decides to buy a new computer to help with homework. In pondering choices, the student considers the advantages of wireless Internet access knowing that the local board of education is considering installing wireless capability in schools. If the student purchases a computer with the extra expense of a wireless connection and the board comes through on its intent, both 'players' can be better off. The student gains the advantage of the Internet connection and the board benefits from having more capable and world-knowledgeable students. However, if the student spends the additional money to buy a wireless enabled computer and the Board decides not to proceed, the student may have wasted money on an accessory that does not help with homework (the original intent of purchase).

EVOLUTION AND ANIMAL BEHAVIOR

Game theory has been very useful when applied to evolution in the traditional sense; that is, the way living things change over time and with environmental pressure. Indeed, in the preface to his 1982 book *Evolution and the Theory of Games*, John Maynard Smith wrote that "[p]aradoxically, it has turned out that game theory is more readily applied to biology than to the field of economic behaviour for which it was originally designed."

The role of game theory in evolution was first developed in 1930 by Ronald Fisher to try to explain the observations that the ratio of males to females in many animals species (he specifically studied mice) are equal, even though the majority of males never mate. However, by using game theory, Fisher deduced that the seemingly 'excess baggage' non-mating males in fact help to raise and protect their grandchildren. Thus, it is in the best interest of the species to maintain fit males to take a role in childcare when females might be aging or in poorer health.

Another aspect of game theory concerns animal behavior. An example of this is the so-called 'Hawk-Dove game.' Animal (yes, even human) behavior can take the role of a hawk; initiating aggression and not backing

down until injured or until the opponent backs down, or of a dove; immediate retreat when danger threatens. When a hawk meets a dove, the hawk will gain the territory, food, nest, etc. When two hawks meet, the feathers will literally fly. When two doves meet, there will be a sharing of the resources or territory. So, depending on who meets whom, the payoff for a hawk and a dove can be rather good or exceedingly bad (e.g., death).

Evolution will seek the path that is the most stable and which carries the most advantage for the species. To return to the hawk-dove example, a dove-dove relationship in a population does not make evolutionary sense, since the presence of even a single hawk throws the system into disarray. In contrast, the aggressive hawk strategy is evolutionarily stable—but if, and only if, the value of the resource in dispute is greater than the cost of being injured in a fight.

Of course, there can be variations on this all-aggression or all-retreat stance. Some animals may chose to fight or cut-and-run, depending on the circumstance. Over time, a dominant strategy (the one that produces the best payoff for the species) will emerge.

Evolutionary game theory need not just be concerned with the Darwinian kind of evolution. Social evolution—the way the beliefs and what is considered to be normal and acceptable behavior a society changes over time—can also be approached using game theory.

Potential Applications

INFECTIOUS DISEASE THERAPY

This application has some overlap to the evolutionary biology area. However, in an era of rising antibiotic resistance by bacteria, and the increasing frequency of infectious disease, the fight against certain infectious bacteria is a stand-alone category.

Game theory has a place in a bacterial population. When a population of millions of bacteria is exposed to an antibiotic, a large proportion of the cells are usually killed. But some will survive, either because they have acquired some genetic or other means of eluding or destroying the effects of the antibiotics, or just because the antibiotic was used up by the time it encountered that individual bacterium.

Game theory has potential in modeling the 'choices' faced by bacteria, and helping guide a strategy that attacks the bacteria after they have made the most probable choice. For example, bacteria exposed to an antibiotic may choose to seek shelter at a surface. By making the surface itself antibacterial, the cells are killed as they colonize the surface. This strategy exploits the bacterial choice as a weapon against infection.

In another tact, researchers at the Center for Genomic Sciences in Pittsburgh, Pennsylvania, and the Center for Interfacial Microbial Engineering at the University of Montana in Bozeman are testing the hypothesis that an individual bacterial cell is part of a larger community, and that the total number of genes in this community can be shared among individual members as needed. Thus, a single cell need not carry every single gene, since it can acquire that gene from its neighbor in time of need.

These sorts of strategies are currently being tested.

EBAY AND THE ONLINE AUCTION WORLD

This potential is real but will surge in coming years. Auction sites such as eBay have a fixed closing time for the auction. Experienced bidders will zoom in with their bid just before the close, often securing the sale. Such "sniping" requires a person to observe the auction, at least periodically, to time their bid. However, there are Internet sites that will automatically assume this function, for a price.

As has been explained by Harvard University economist, Al Roth, sniping web sites and the experienced auction snipers are examples of game theory in action. Late bidders are less influenced by the decisions of others. As well, those who recognize the value of an item will often reserve their bid for the end of the auction so as not to tip their hand about the object's value, thus hopefully keeping the price down.

Knowledge of game theory could make for more successful Internet auction hunting.

ARTIFICIAL INTELLIGENCE

A computer is not yet the equal of a human. Humans can make an independent decision, but computers must be pre-programmed with decisions that are based on the meeting of a number of conditions. Computers cannot make decisions they have not been programmed to make.

In the future world of artificial intelligence, it is anticipated that computers will be able to make decisions free of the constraints of meeting pre-set conditions. Game theory can play a role in this new world. Artificial intelligence programs would be able to make new decisions that produced the best payoff, based on the incoming information and on the previous experiences of the computer. In other words, a computer would be capable of "learning."

Key Terms

Nash equilibrium: A set of strategies, named after John Nash, that results in the maximum benefit of each player.

Player: In game theory, a decision maker.

Plays: In game theory, choices that can be made.

Zero-sum game: An outcome of a game where players choices have produced neither a win or a draw for all of the players.

Where to Learn More

Books

Camerer, C.F. *Behavioral Game Theory: Experiments in Strategic Interaction.* Princeton: Princeton University Press, February 2003.

Fisher, T.C.G., and R.G. Waschik. *Managerial Economics: A Game Theoretic Approach.* New York: Routledge, August 2002.

Miller J.D. *Game Theory at Work" How to Use Game Theory to Outthink and Outmaneuver Your Competition.*" New York: McGraw-Hill Trade, March 2003.

Osborne M.J. *An Introduction to Game Theory*. Oxford: Oxford University Press, August 2003.

Periodicals

Greig D., and M. Travisano. "The Prisoner's Dilemma and polymorphism in yeast SUC genes," *Proc R Soc Lond B Biol Sci.* (February 2004) 271: S25–S26.

Holt C.A., and A.E. Roth. "The Nash equilibrium: a perspective," *Proc Natl Acad Sci USA* (March 2004) 101: 3999–4002.

Kirkup B.C., and M.A. Riley. "Antibiotic-mediated antagonism leads to a bacterial game of rock-paper-scissors in vivo," *Nature* (March 2004) 428: 412–414.

McNamara J.M., Z. Barta, and A.I. Houston. "Variation in behaviour promotes cooperation in the Prisoner's Dilemma game," *Nature* (April 2004) 428: 745–748.

Taylor P.D., and T. Day. "Behavioural evolution: cooperate with thy neighbour?" *Nature* (April 2004) 428: 611–612.

Web sites

Alexander, J.M. "Evolutionary Game Theory." <http://plato.stanford.edu/emtries/game-evolutionary> (May 27, 2004).

King W. "Game Theory: An Introductory Sketch." <http://william-king.drexel.edu/top/eco/game/intro.html> (May 27, 2004).

Levine DK. "What is Game Theory?" <http://levine.sscnet.ucla.edu/general/whatis.htm> (May 27, 2004).

McCain, Roger. "Essential Principles of Economics: A Hypermedia Text" <http://william-king.www.drexel.edu/top/prin/txt/EcoToC.html> (Oct 23, 2004).

Geometry

Overview

Geometry is defined as the branch of mathematics concerned with the properties and relations of points, lines, surfaces, and solids, and can be found in absolutely anything, from the shape of the tires on a car, to the movement of light throughout the Universe. The forms of buildings, the paths of flying objects, the patterns in strands of DNA, the characteristics of leaves on a tree, the markings on a sports field, and the behavior of atoms can all be investigated using some degree of geometric reasoning. Even someone who has never heard of the mathematical study of geometry cannot help but use its concepts in every action and observation. In its entirety, geometry relates all objects and phenomena to the mathematical language common to all human communities on Earth, existing at the heart of all indisputable logic.

Fundamental Mathematical Concepts and Terms

In order to understand how geometry can be used to expound upon just about anything in the Universe, it is important to fully grasp the concept of dimensions. Although all objects in the Universe have three dimensions, they can all be related to lesser dimensions.

Theoretically, points have no dimension. A point is a set of coordinates, a location, not something tangible. When a point is represented as a drawing on graph paper, a dot is used to represent the location, but the mark that is used to indicate the location of a point is always three-dimensional. For example, a dot made by a pencil on a piece of paper is roughly circular, having a crude diameter, perimeter, and area existing in two dimensions, and a measurable thickness, or height protruding from the surface of the paper to account for the third dimension.

To represent one dimension, a pencil can be dragged across a piece of paper to draw a curve (where a line is considered to be a curve with no curvature). This drawing also exists in three dimensions, having thickness, width, and length. However, the drawing is intended to represent only one of these dimensions, length. Lines are considered to continue forever along their path, while line segments include only the section of the line between two endpoints. When two straight lines or line segments meet, they create an angle. Two lines are said to be parallel if they will never meet. Perpendicular lines create a 90° angle, or right angle, like those found in rectangles. Angles smaller than 90° are called acute angles; larger angles are called obtuse angles.

The geometric alignment of the stones in Stonehenge served as an ancient astronomical calculator and calendar. PHOTOGRAPH BY JASON HAWKES. CORBIS. REPRODUCED BY PERMISSION.

When one-dimensional curves create an enclosed shape, the realm of two dimensions is reached. Two-dimensional figures exist in planes, which can be thought of as pieces of paper for drawing shapes, although planes have no thickness and exist only in the imagination. Polygons and ellipses are fundamental examples of two-dimensional figures. Any enclosed planar figure (existing in a plane) bound by straight lines is defined as a polygon. Basic polygons include triangles, quadrilaterals such as parallelograms and rectangles, pentagons, hexagons, and octagons. Because of their many interesting properties regarding lengths and angles, triangles create a large amount of useful concepts. The study of triangles is referred to as trigonometry and is most often taught in a separate academic course.

Ellipses include ovals and circles, and enclose two-dimensional space without creating any corners. The definition of an ellipse involves two points, but the concept is easier to understand by considering the definition of a circle. A circle is a special ellipse that has a boundary defined as the set of points that are equidistance (the same distance) from one point. That distance is called the radius and is usually represented by a line segment stretched from the center to the perimeter. The perimeter can be viewed as the path that the end of the radial line segment traces out when is rotated around the center like the hands of a clock. In any other ellipse, there are two points that determine the boundary, causing the figure to resemble a circle that has been stretched. Ellipses are always symmetrical, meaning that if they are cut in half at some angle, each side looks like a mirror image of the other side.

Vectors present another important geometric concept, representing both a number, known as a vector's magnitude or length, and an angle that determines its direction. Vectors are essential to depicting phenomena such as wind or the movement of a sailboat in water. When investigating the factors that determine the movement of a sailboat, the vectors representing the speeds and

With its five-sided geometric shape still intact in the inner rings, the west face of the Pentagon is shown in this aerial photo taken before the attack on September 11, 2001. UPI/CORBIS-BETTMANN. REPRODUCED BY PERMISSION.

directions of the wind and water currents can be added to the vector representing the motion of the vessel in order to model the true motion of the boat relative the seabed.

Any geometric figure can theoretically exist in three dimensions. For example, the point 5 feet (1.5 m) above the intersection of the borders between Utah, Colorado, Arizona, and New Mexico can be described in reference to other locations on Earth, but that point is not a tangible object and has no dimension. The border between Utah and Arizona can be described in reference to other geographic landmarks, but it has only one dimension that describes the imaginary path of the border. The border around Colorado can also be defined as a rectangle with a definite length and width, but this two-dimensional figure cannot be seen or touched. Someone standing at the intersection of the four states can imagine all of these figures, and even though they do not actually exist, their underlying concepts are vital to our perception and use of space.

In reality, the form of every object has three dimensions. The notions of lesser dimensions, however, provide helpful tools for understanding all that is real. Humans can simplify any physical problem by investigating the properties of points, curves, and areas. For example, the area of a circle is defined by $A = \pi r^2$, where A represents the area and r represents the radius; and the formula for the volume of a cylinder is $V = h(\pi r^2)$, where V represents the volume and h represents the height. It is easier to understand the formula for volume by thinking of a cylinder as a circular base extended upward. The volume is equal to the area of the two-dimensional base multiplied by height, the third dimension of the cylinder.

In the mind of anyone equipped with basic geometric tools, anything in the physical universe can be transformed into combinations of geometric figures spanning multiple dimensions. Whether too small to see, plainly visible, or located in a different solar system, any object

seems less intimidating when represented by elegant geometric figures that have been systematically scrutinized for thousands of years.

A Brief History of Discovery and Development

Like most fields of mathematics, geometry found its beginning due to a necessity to understand and predict phenomena in the natural world. Geometry provided a tool to help humans understand their surroundings long before generalized mathematical formulas were conceived. As long as 6,000 years ago, people began using geometric reasoning as a visual aid to explain and predict phenomena in the world around them. Pure mathematical reasoning (and proofs that the reasoning was sound) would not appear until geometry was coupled with algebraic rules thousands of years later.

Possibly the first use of abstract geometric reasoning was invoked in early human settlements as result of the inception of monetary calculations. In Egypt, for example, nomadic peoples led simple yet inconsistent and insecure lives, roaming the land and setting up temporary shelters wherever the forces of nature (most notably weather and the availability of food) guided them. After thousands of years spent roaming the sparse desert, these people eventually discovered that the Nile River provided ample water, and hence vegetation, which in turn attracted animals, essential survival resources that are consistently sparse in the middle of the desert. The activities of their lives depended on the ebb and flow of the river. Even their first calendar was based on the cycles that governed their annual cultivating and harvesting periods. After generations of cultivation and expansion along the river, towns were set up and a system of leadership was established.

The concept of money was implemented in these towns in order to facilitate trades between the citizens, and eventually visitors. Similar to the taxation of modern cultures that enables important communal facilities, including schools and healthcare, the Egyptian leaders set up a system of taxation to support the needs of the growing culture. Most of the taxes implemented under the Pharaohs were used to build large, intricate structures to support their images as divine beings that should be feared and followed. Because the Nile held the key to the flourishing agricultural settlements, the River was believed to possess a godlike persona, and each year the taxes were based on the amount of flooding that took place. Taxes were vital to the success of the Pharaohs, and were therefore taken very seriously. Arithmetic and algebra (the essential tools for analyzing numbers) had yet to be discovered, so Egyptian accountants used visual aids, now recognized as geometric figures, to determine the amount that each citizen should be taxed. The circle, for one, was an important figure in the calculation of taxes. This use of circles led Egyptians to approximate the value now commonly symbolized by the Greek letter π, which defines the relationship between radius and circumference. They estimated pi (π) to be 3.16, but it is closer to 3.1415927. The fact that Egyptian officials were happy to collect slightly too much tax from each citizen is probably the reason that they used an overestimate of the value for π, when an underestimate could surely be calculated just as easily. In a culture that called for the execution of mildly rebellious individuals, a percentage of income was not something to be quibbled over.

In ancient Egyptian culture, spatial measurements were eventually given a name that translates loosely to "Earth measure." The Egyptians and Babylonians used these ideas to describe the physical world around them, but made no advances in using mathematics to explain seemingly inexplicable events or reveal fundamental truths. The building of the pyramids is the most notable use of geometric reasoning by ancient Egyptians. A pyramid has a square base and four triangular sides that come to a point at the top. If a measurement is even slightly inaccurate, the top of the pyramid will not come to a point, but rather a flat grouping of stones. This would subtract from the pyramid's magnificence and almost certainly result in the execution of the designer at the order of the Pharaohs who demanded divine perfection.

To ensure that the base was square and that the triangular faces met at a point, the Egyptians conceived an ingenious method of measurement. Systems of ropes were stretched by a handful of workers to map out triangles. Knots tied in the ropes at equal distances enabled the workers to create triangles with proportional lengths, in turn controlling the resulting angles. For example, in a triangle with a side created by three knots, a side created by four knots, and a side created by five knots, the angle opposite the longest side will be a right angle.

With all of their advances in measurement, the Egyptians did not view the concepts behind their quantitative reasoning as mathematical notions that could be extended to explain other phenomena. They did not truly understand the theoretical points represented by the knots, or the lines represented by the ropes.

During approximately the same time period that gave rise the Egyptian pyramids, the Babylonians were making similar advancements in the perception of geometric logic.

In general, the civilizations that existed prior to the Greeks noticed and utilized many interesting and helpful properties about their physical surroundings, and were able to use their findings to accomplish relatively incredible feats. What separates these earlier cultures from the Greeks is that they did not possess, and perhaps did not desire, a deeper understanding of the measurements that they recorded.

About 2,500 years ago, a Greek merchant, engineer, and philosopher named Thales at long last expanded geometric reasoning beyond the measurements of taxes and sand, sparking a new wave of logic and abstraction that would later motivate the work of Pythagoras. Thales invested most of his life to the evolution of all types of knowledge, including astronomy. For part of his life, Thales lived and studied in Egypt. It is widely accepted that he was first to determine the true height of one of the pyramids, employing the triangles used in the precise measurement of the base in a completely new way. When he returned to Greece, he brought a multitude of new ideas regarding spatial measurements. These ideas and the resulting field of logical reasoning that followed were soon thereafter referred to as geometry, the term that is still used today. Like the term used in Egypt at the time, the word geometry translates loosely to "Earth measurements."

Ancient sources claim that, shortly before death, Thales received a visit from a young philosopher and mathematician named Pythagoras. During their meeting, Thales is believed to have suggested that Pythagoras travel to Egypt in order to further the advancement of Egyptian practical geometric concepts. Pythagoras followed this advice, returning to Greece with another generation of knowledge pertaining to mathematics, astronomy, and philosophy. For example, though it is evident that the Pythagorean theorem was used in the creation of pyramids thousands of years earlier, it is named after him due to the growing Greek affinity for logic. No one in Egypt would have been named after such a concept, because it was not regarded in that land as important knowledge outside of the occasional launch of a new pyramid. Pythagoras found many followers in Greece, and eventually settled with them outside of Athens to live life truly by numbers. To the Pythagoreans, everything in life was numerical. Mathematics was essentially their religion. Outsiders were generally regarded as blasphemers, unworthy of the beautiful truths of advanced mathematics. Their notion of irrational numbers (numbers that cannot be represented as one whole number divided by another) was one of their greatest secrets, and one member was allegedly drowned for leaking this knowledge.

About two centuries later, Euclid (c. 325–265 B.C.), another Greek mathematician, again enhanced the study of geometry. While little of his work provided any original ideas, Euclid is often regarded as the father of geometry because his systematic methods for representing geometric ideas sculpted the subject into a more manageable form. Using consistent and simple notation, geometry could be studied by all and generalized to fit more and more situations. Euclid's work has provided the basis for communicating geometric ideas ever since.

With the rise of the Roman Empire came a heavy decline in intellectual advancement. During the ensuing 900-year period referred to as the Dark Ages, the works of Euclid and his predecessors were locked up in private libraries, and not until the seventeenth century would geometric thought continue to advance.

In the early seventeenth century, an Italian philosopher and mathematician named René Descartes revolutionized both fields in ways that have yet to be improved upon. In mathematics, he melded the laws of numbers and geometric measurements by conceiving the coordinate plane. By placing geometric figures in a well-labeled grid, algebraic manipulations could be applied directly to geometric figures. In this way, geometric concepts began to be analyzed in new ways, illuminating the concepts of the past and spawning an enormous amount of new theories. The simultaneous consideration of algebraic and geometric properties first introduced by Descartes has come to be known as analytical geometry, and has provided the basis for all scientific endeavors in the four centuries since his life.

In modern times, scientists have realized that the three space dimensions are not truly independent of the time dimension. This concept was popularized by the German-born American physicist and mathematician Albert Einstein (1879–1955) in the early twentieth century. Just as the works of Thales, Pythagoras, Euclid, and Descartes were difficult to grasp when they were first introduced, the idea that time and space interact complicates geometry to a degree that most people are not yet equipped to understand. However, as time advances, so shall the collective human understanding of it as a concrete dimension.

Real-life Applications

POTHOLE COVERS

Millions of potholes, also called manholes, are scattered throughout the world, giving workers access to underground sewer lines. In the United States, these potholes and the large hunks of metal that cover them are almost always round. A pothole cover is a cylinder with a

relatively small height. This shape is chosen because of two important properties about circles.

First of all, a circle is one of only two basic two-dimensional geometric figures that cannot fit through an identical figure in three dimensions. That is, if two identical circles (having radii of equal length) are placed in three-dimensional space, one could not be made to pass through the other, no matter how the circles were angled. All but one other basic shape can be lifted and rotated so that it will fit through a copy of itself. A square, for example, can be rotated upward until vertical, and then rotated approximately 45° in any other direction to fit diagonally through a copy of itself because the diagonal of a square is longer than any of its sides (a fact which can be proven using the Pythagorean Theorem). Similar reasoning can be used to show that almost all other basic shapes can be made to fit through an identical shape. The fact that a circle will not slip through an identical circle ensures that heavy round pothole covers do not fall through the openings of potholes and injure the workers below.

An equilateral triangle (a triangle with three sides of equal length and three 60° angles) is the only other basic two-dimensional geometric figure that cannot be made to pass through a copy of itself. If the lengths of the sides of the two identical triangles were not all equal, one of the triangles could be propped up vertically with a side other than the longest on the bottom, and then rotated until the bottom side of the vertical triangle was parallel to a longer side of the horizontal triangle, and fall through.

In some parts of the world, pothole covers are occasionally found in the form of an equilateral triangle and they, of course, never fall through the hole. However, most pothole covers are chosen to be round because of another unique characteristic of circles; they have no corners. A circular pothole cover saves the energy of workers by allowing them to roll the heavy covers out of the way. The three vertices of a triangle also make it more likely to injure someone while it is being moved or sitting on the ground. A circular shape also makes replacing pothole covers easier because they do not need to be rotated in order to line up with the shape of the hole.

ARCHITECTURE

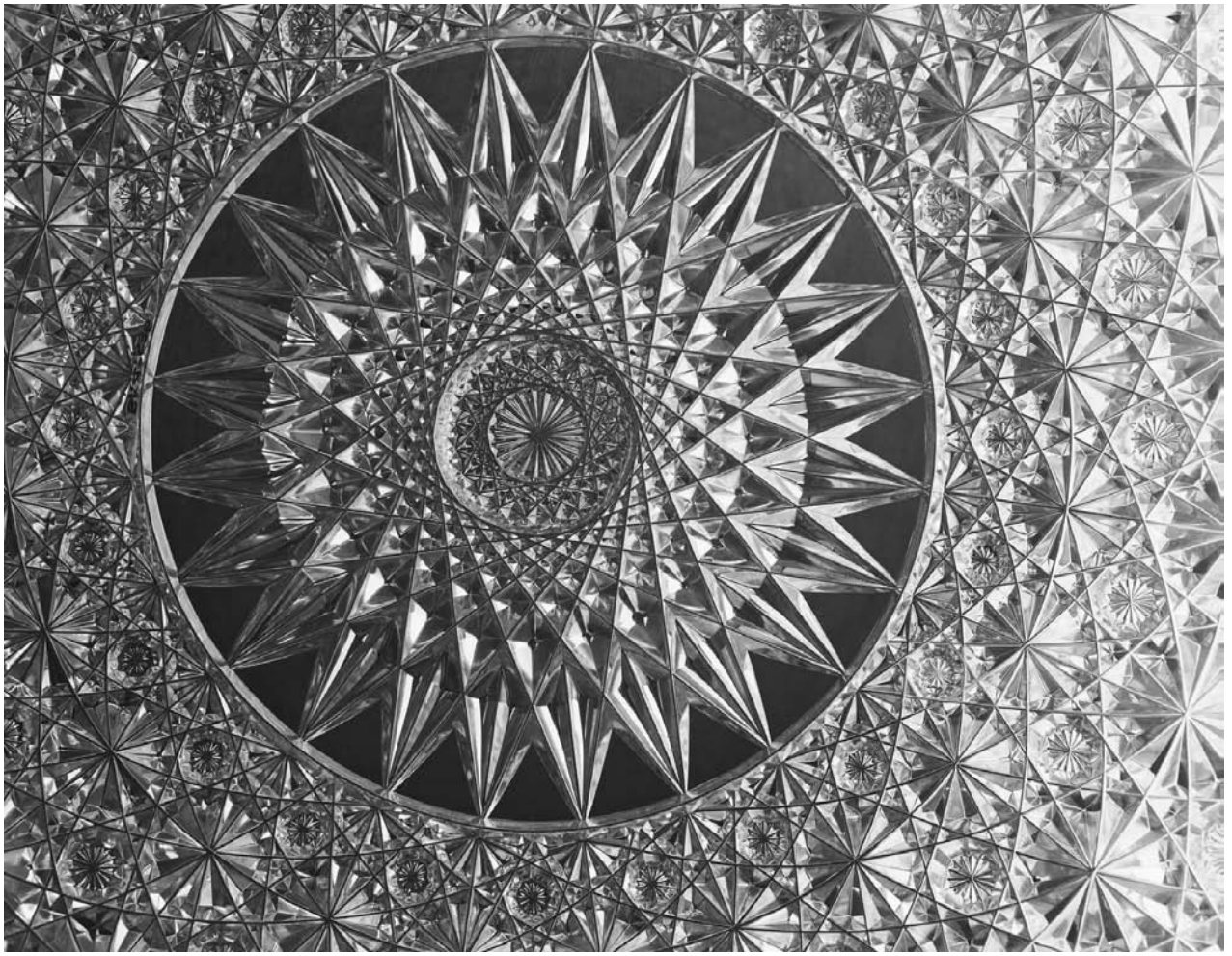
Mathematical reasoning was used to build the Egyptian pyramids over 6,000 thousand years ago, long before geometry was conceived as a fundamental field of mathematics. Since the rise of the early Greek mathematicians, standardized geometric concepts have provided the essential tools for planning and constructing all types of buildings.

All architectural structures—from simple four-walled buildings with flat roofs to elaborate, multipurpose constructions—comprise combinations of geometric figures. The types of curves and shapes used to create an enclosed space are carefully chosen for their effects on function and beauty.

In the design of some buildings, functionality is far more important than artistic expression. For example, when building a silo for the storage of large amounts of grain, a cylinder provides the most efficient use of space. The horizontal cross-section of a cylinder (e.g., the base) is circular, and circles can be shown to have the largest area with respect to perimeter out of all two-dimensional shapes. Because the circular cross-section is extended vertically to form a three-dimensional cylinder, a cylinder requires the least surface are—and therefore the least building materials—to provide a given volume.

Most temples and churches, for instance, are designed to balance the beauty required for paying respect to the religious figures worshiped by the congregation (e.g., painted ceilings, intricate moldings, and stained picture windows) with the function of fitting a large amount of people into a single enclosed space. When congregations first began to grow, the task of building a room of worship large enough to accommodate everyone posed a serious problem. Enclosures had previously been created using only flat walls and ceilings, which greatly limited the space because the flat ceiling would collapse if the supporting walls were too far apart. This obstacle was eventually overcome by the Romans when they began using semicircular arches in their architecture. A semicircular arch is created by stacking half of a circle on top of a rectangle. This structure allows the supporting walls to be further apart because, as gravity pulls on the structure, the semicircle distributes weight so that the arch does not collapse. Soon thereafter, arches were placed one in front of the other to create the walls and roofs of relatively huge halls, giving congregations room to grow. Domes are created by copying and rotating an arch around its highest point and were later added to create even higher ceilings, instilling a sense of awe worthy of a building devoted to gods. The round nature of domes also adds a sense of perfection, an important aspect in many religions. Individual arches were also used as the shapes of windows and doors to increase structural stability and effect beauty.

The arch is just one of the geometric figures integral in the design of various structures. Like an arch, a triangle adds stability by distributing and balancing the downward force of gravity as well as the lateral forces of wind. The triangular configuration of rafters in the roofs of



Intricate geometry contained in glasswork. BETTMAN/CORBIS.

many houses also allows rain and debris to slide down and fall off so that the roof does not have to endure much weight. Sloping roofs also add a familiar aesthetic characteristic to the house's overall appearance.

Rectangles, ellipses, hexagons, and more can be found throughout any human construction, including the landscaping that surrounds many buildings. Most structures made by humans are heavily dependent on geometric figures, from the triangles found in the enormous towers that support elaborate networks of electrical wiring to the visually pleasing dimensions of national monuments. The 630-foot (192-m) tall Gateway Arch in St. Louis provides another example of the abilities of the geometric arch to create a sense of beauty while stretching the limits of architecture. The Washington Monument consists of four trapezoidal sides topped by a pyramid. This gigantic vertical structure is impressive for its size: the square base of the shaft has a width of 55 feet (16.7 m); the shaft is 500 feet (152 m) tall; and the pyramid is 55 feet tall. Its construction required in-depth analysis of lengths, angles, areas, and volumes.

Some structures, such as the pointy onion-shaped roofs common in the regions of the former Soviet Union, do not appear to be associated with basic geometric shapes. Other buildings, including some modern museums, are intentionally designed to appear irregular in shape, seemingly not following any laws of mathematics. However, all of these shapes are defined and thoroughly analyzed using mathematical formulas describing complicated geometric figures in one, two, and three dimensions.

Regardless of the goals of the architect, every aspect of a building is deeply considered. Every consideration—from the orientation of the wood panels on a residential property to the ponderous calculations involved in building a retractable dome over a gigantic sports

The Washington Monument consists of four trapezoidal sides topped by a pyramid. CRAIG AURNESS/CORBIS. REPRODUCED BY CORBIS CORPORATION.

stadium—are modeled and investigated using geometric reasoning.

HONEYCOMBS

Just like geometric figures are used to affect certain functions in a structure, many species of animals seem to utilize a small amount of instinctive geometric knowledge in order to conserve materials and energy. Various species of spiders, for example, spin webs in patterns that maximize the ability to catch bugs while minimizing the amount of silk expended. Birds generally create circular nests because a circle provides the maximum amount of area for a given amount of materials.

Honeybees also display an instinctive understanding of geometry in their use of hexagonal chambers for storing honey in honeycombs. One might think that bees would choose to build chambers with circular holes (cylinders) in their honeycombs because, just as a circle allows for the most area given a specified perimeter, a cylinder provides the largest volume for a given perimeter. However, placing a group of circles next to each other wastes space because they do not fit together. For example, if a group of cylindrical silos were built side by side and viewed from above, it would be easy to see that space is wasted between the silos where grain cannot be stored. Most two-dimensional geometric figures cannot be used to completely fill a two-dimensional area. In fact, only equilateral triangles, rectangles, and hexagons can fit together with identical figures to completely fill an area.

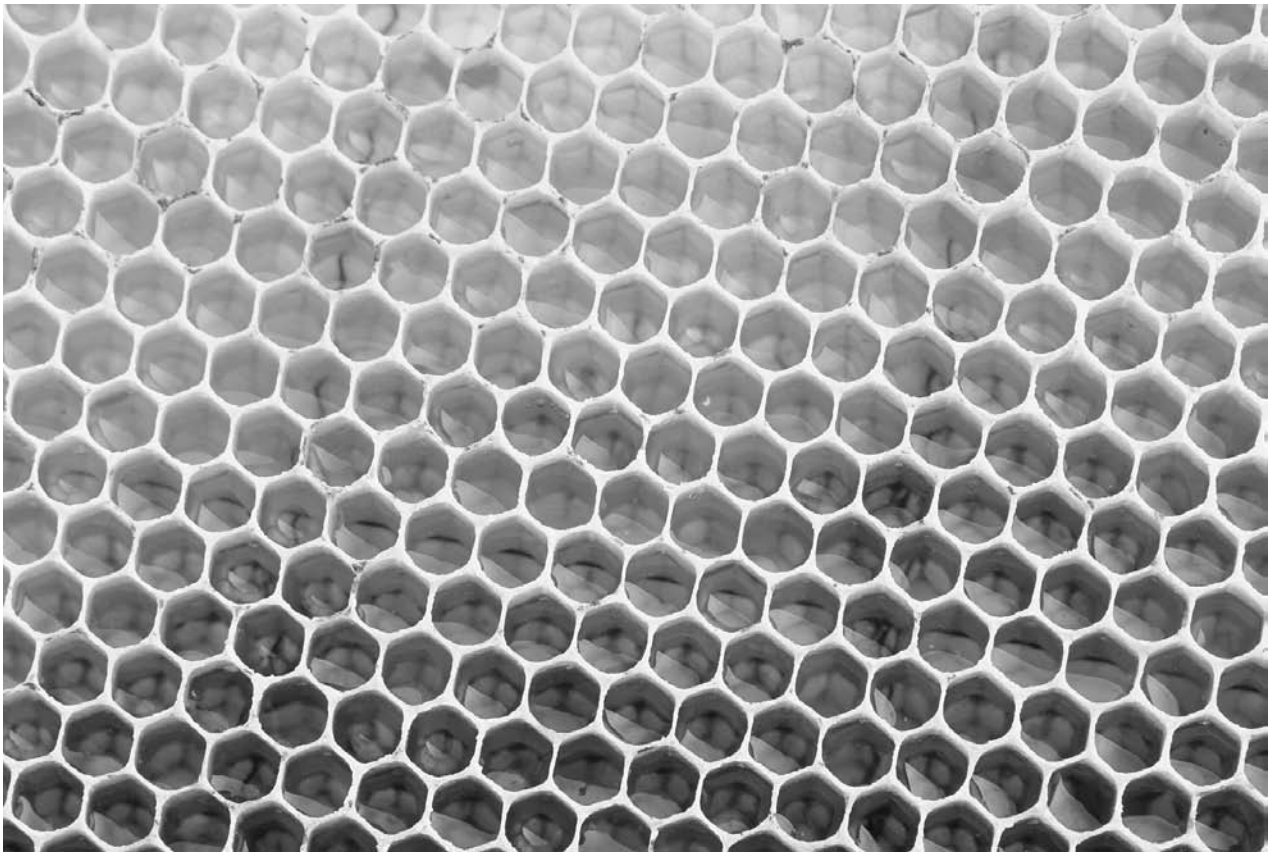
Of these three space-filling shapes, the hexagon is the most efficient—that is, it uses the least material (wax, in this case) to make a honeycomb having a given volume. It might seem strange to think of an insect using geometric reasoning but over time nature has a knack for uncovering the most logical solution to any problem.

GLOBAL POSITIONING

In the new millennium, a global positioning system (GPS) receiver is all a person needs to determine his or her exact location on Earth. Every day, thousands of campers, hikers, bikers, skiers, hunters, boaters, pilots, and motorists around the globe use these ingenious devices to ensure that they do not get lost.

The global positioning system consists of 27 solar-powered satellites that orbit Earth. Only 24 of the satellites are in operation at all times; the extra three are in orbit in case any of the operational satellites malfunctions. This network of satellites was originally launched by the United States government to aid in military navigation, but was shortly thereafter made available for use by anyone. Each satellite weighs about 4,000 pounds (1,814 kg), orbits at about 12,000 miles (19,312 km) above the surface of Earth, and completes a rotation around Earth approximately twice a day. From any position on the surface of Earth, at least four of these satellites can be detected by a GPS receiver at all times.

In order to understand how a GPS receiver uses satellites to determine its own location, it is helpful to first discuss the process in two dimensions. Imagine, for example, that an explorer is lost somewhere on Earth with a detailed map, but absolutely no idea of her current location. She asks one of the locals if he can help, but the only information she receives is that she is currently 699 miles (1,125 km) from Barcelona. Because she does not know the direction in which Barcelona lies, this only indicates that she is somewhere on the perimeter of a circle with its center in Barcelona and a radius of 699 miles. This is a good start, so she draws this circle on her map. Hoping to



Hexagons in honeycombs. RALPH A. CLEVENGER/CORBIS.

get a better idea of her location, she asks another person for help. This time, she learns that she is currently 680 miles (1,094 km) from Berlin. Again, this means that she could be at any point on a circle around Berlin. After drawing this circle on the map, she notices that there are now two intersecting circles on which she might currently be located. For both of the pieces of information she received to be true, she must be located at one of the two points where the circles intersect. Having two reference points lowers her possible locations from an infinite number of points around a large circle down to only two points. To figure out which of these two points represents her location, she needs one more reference point, which will generate another circle of possible points. The next person she talks to tells her that she is currently 257 miles (413.6 km) from London. She draws a circle around London with a radius of 257 miles and finds that it only intersects one of the two possible points found with the first two circles. This point represents her current location, which turns out to be Paris.

GPS receivers use similar reasoning extended to three dimensions. By measuring the lag in radio waves sent from a satellite, the GPS receiver determines how far the satellite currently is from the receiver; but the direction from which the radio waves approached is not determined. Just like a circle can be staked out on a map, the distance to a satellite allows the GPS receiver to map out a sphere in space. The GPS receiver must be located somewhere on the boundary of this sphere.

Next, the GPS receiver locates another satellite and determines how far away it is, creating another sphere. The intersection of the two spheres is a circle. (Imagine chopping off a thin portion of two oranges to create flat surfaces so that they can be stuck together; the boundary of the area where the two oranges touch is a circle.) Using these two spheres allows the GPS receiver to narrow its location down to a circle, just like using a single reference point on a two-dimensional map creates a circle. Because one more dimension is involved, it takes one more reference point to narrow the location down to a two-dimensional circle.

The GPS receiver then locates a third satellite, mapping out a third sphere. Just as a second circle drawn on a two-dimensional map determines two possible points, the

third sphere determines two points in three-dimensional space. At this point, the receiver can approximate its location by using Earth as a fourth sphere. One of the two points created by the first three spheres is always out in space, so the GPS receiver knows that the point that lies on Earth's surface is its approximate location.

Errors can occur in the calculations performed by GPS receivers, due mainly to the effects that Earth's atmosphere has on the speed of the radio waves sent out from the satellites, which throws off the perceived distance to each satellite by a small amount. So, while three satellites are theoretically enough to pinpoint the location of a GPS receiver, these devices always attempt to locate a fourth satellite in order to increase the accuracy of their calculations.

FIREWORKS

Fireworks employ two different types of explosive powder, flash powder and black powder. Flash powder is used to create bright a bluish-white light that can easily be changed to any color by adding certain chemicals. Almost all fireworks that create light include flash powder. Sparklers, for instance, consist of flash powder stuck to a small metal rod. Black powder (also known as gunpowder) is usually used in fireworks to create a loud explosion. Basic firecrackers contain mostly black powder. Some of the larger and more elaborate fireworks available at firework stands near the 4th of July consist of a cylindrical casing that houses combinations of flash powder and black powder. The black powder creates noise and launches projectiles into the air, while the flash powder emits bright light from the main casing and airborne projectiles. Projectile charges that create light during their flight are called flash star pellets, or stars.

Large public gatherings at sports stadiums and other outdoor venues feature enormous displays of fireworks, creating explosions that are audible for miles and visually stunning color patterns that can seem to take up the entire sky. Different patterns of light (e.g., spheres and discs) are achieved by mapping out the desired pattern of the stars inside of the large casing. Most main casings are cylindrical, with the height and size of the base dependent on the intended size and shape of the color pattern. The stars are separated and held in place by black powder, which fills the remaining volume of the casing. To cause a spherical pattern, the stars are equally spaced throughout the three dimensions of their casing. To create a disc that spreads in a plane parallel to the locally flat surface of Earth, the stars are situated in a circular pattern around the center of the casing; if the disc sits at an angle inside of the casing, a similar angle will appear in the explosion.

Spheres and discs are the most logical shapes because they require that the stars are all projected the same distance. Those distances, the radii of the spheres and discs, are determined by the amount of black powder packed between the stars. Therefore, a larger casing usually results in a larger spread of stars because there is more extra room for black powder.

In order to launch the whole thing into the sky, the firework is first placed into a cylindrical tube with black powder in the bottom. The correct volume of black powder is determined based on the mass of the firework and the desired height to be reached. The black powder in the tube is then ignited, shooting the firework upward (and possibly at an angle determined by the angle of the tube) and igniting the firework's main fuse. All of the fuses in the contraption have a pre-calculated length in order to control the timing of the explosions of light and sound.

MANIPULATING SOUND

Using the right materials and a little geometric reasoning, enclosed spaces can be designed to affect sounds in different ways. The muffler attached to the exhaust pipe on a car, for example, is intended to absorb most of the noise produced by the engine, but the thin materials used absorb a relatively small amount of the noise. The majority of the engine noise is canceled out by additional waves of sound produced when the engine's sound waves bounces off the walls and inner structure. Sound from the engine enters the muffler through a pipe, where some of the sound waves escape through circular holes in the side of the pipe, called perforations. The dimensions of the pipe and the size and position of perforations are important to directing the sound against different walls of the muffler. The main part of the muffler is sectioned off into three chambers, and as sound waves bounce in different directions, they cancel each other out. Technically, the frequencies of the sound waves interfere with each other, and the combination of sound waves results in fewer vibrations in the air. Because a muffler cannot consist of, say, padded concrete walls, it must be designed to use sound waves against each other in order to disrupt as much of the noise as possible.

A concert hall, on the other hand, is designed to bounce sound waves around a large volume in order to enhance the sound. The dimensions of the walls, ceiling, and floors surrounding the performers and audience greatly affect the resulting sounds, as do the furniture and anything protruding from the walls or ceiling. The most important consideration when designing a concert hall is the equalization of the different frequencies coming from

the stage. The sound waves that produce higher sounding notes do not tend to bounce of surfaces as well as the slower vibrations of lower sounding notes. The overall shape of the hall and the additional surfaces (especially shapes and patterns incorporated into the walls and ceiling) are intended to ensure that the best balance of tones makes it to each listener's ears.

Reverberation is another important aspect of an acoustic space. Reverberation relates to the amount of time that sound waves continue bouncing off of the various surfaces. As a simple interpretation, reverberation defines the amount of echoing effect that will be heard. A mathematical formula, called the Sabine formula, can be used to measure the relative amount of reverberation and to determine how much absorbing surfaces need to be added to or removed from a space to achieve optimal reverberation time.

Acoustic instruments rely on the same types of considerations on a much smaller scale. The sounds produced by any drum that does not rely on electricity depend almost entirely on its shape and size. The dimensions of a piano can greatly affect the quality of sound produced when the keys are pressed. On an acoustic guitar, the strings are attached to the outside of the wooden body; but the dimensions of the body are the main reason that the sounds produced by the strings can be heard clearly. Much of the vibrations coming from the strings enter the body of the guitar through a circular sound hole situated directly beneath the portion of the strings that is most often plucked or strummed. The body is carefully designed so that the entering sound waves are most likely to bounce at angles that enhance the sound in ways similar to sound waves bouncing around a well-designed concert hall.

The geometric dimensions of strings are a major factor in the tones produced by any stringed instrument. The strings are essentially cylindrical, having a small circular cross-section stretched along a third dimension to create a volume. On an acoustic guitar each string is stretched to the same length, so the sounds produced by the different strings are dependent on the area of the circular cross-section (the thickness). A larger cross-sectional area results in a larger volume. Because the strings are made of similar materials, thicker strings also have a greater mass, which causes them to vibrate slower and to produce lower sounds than thinner strings.

Reverting to the single dimension of curves and lines, guitar strings can be thought of as parallel line segments. Strips of metal (called frets) are spaced out along a wooden fret board, or fingerboard. The frets lie perpendicular to the strings, defining important points where they intersect. Pressing a string against one of the frets essentially changes the length of the segment being strummed or plucked, causing the strings to vibrate at different frequencies and altering the resulting tones.

SOLAR SYSTEMS

All things in the physical universe, from molecules to exploding stars, have forms that can be defined geometrically. The laws of the Universe have worked together over the past few billion years to create incredible geometric shapes.

For example, solar systems all over the Universe tend to be relatively planar, like huge spinning discs in space. Basically, as a star begins to be crushed by gravity, pulling in all sorts of nearby materials, it picks up a spinning motion, similar to the spinning motion of tornadoes or water escaping through a drain in a bathtub. (If gravitational shrinkage keeps up for too long, a black hole results.) Things continue to spin, and like a ball of dough spun in the air to create a flat pizza crust, the ensuing solar system expands to a practically flat shape. Because of the relative emptiness of space, things spread out at a somewhat constant speed, creating an elliptical disc expanding in a plane. As the rocks, gas, dust, and other debris spin around the star, they collide and collect together to form planets, moons, comets, and asteroids.

The planets are the largest collections of materials and continue on an elliptical orbit around the star. Because they are so big, planets create their own substantial amount of gravity and attract debris that settles into orbit around them. Sometimes this debris collides, and eventually creates a single satellite around the planet. Planets and moons are suspended in an orbit, so each collision causes the material to spin (similar to the effect of flicking a coin held vertical by one finger) eventually leading to large bodies spinning on a constant axis and taking spherical forms. When the debris collecting around a planet has not collected to form a single satellite, the planets are encircled by planar belts of debris.

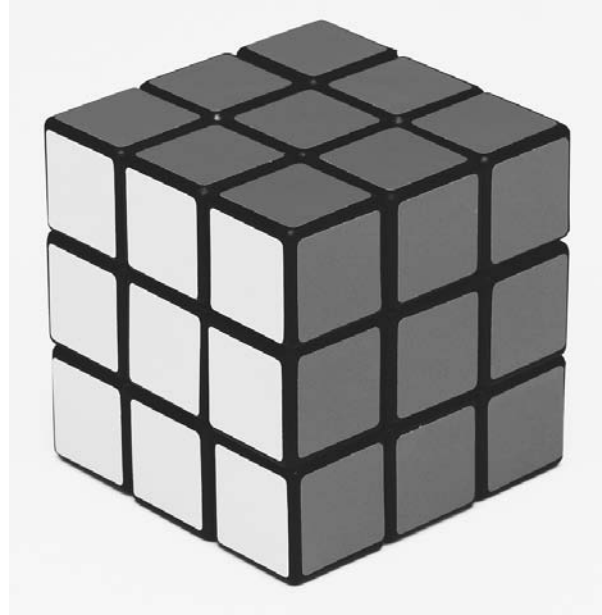
As new debris enters the atmosphere, it is attracted to the belt by the spinning forces. It is no coincidence that all of these spinning objects take on elliptical or spherical forms. These heavenly bodies all provide stellar examples of the idea of a radius defining a collection of points equally spaced from a center.

As moons spin around planets that are spinning around a star, all of the orbiting bodies create thousands of constantly changing angles between them. These changes in these angles are rather periodic, and are often studied and accurately predicted from Earth. For example,

lunar and solar eclipses, in which Earth and the moon line up to create a straight line with the sun, are marked on many standard calendars.

Until late in the twentieth century, it was thought that Earth's solar system might be the only true solar system in the known Universe. Because most stars are unfathomable distances from Earth, their intense light drowns out any nearby material, no matter how powerful the telescope. However, as planets orbit around the star, they cause it to move around in a relatively small circle. From a viewpoint on Earth, this movement appears as a minute back-and-forth motion. Using mainly the star's wobble as an indicator, astronomers can determine the number of planets, the mass of each planet, and their relative distances from the central star.

Scattered throughout the Universe and confined to planes tilted at various angles, brilliant solar systems and exploding star systems (that may or may not become solar system sometime in the vast future) illustrate nature's affinity for geometric figures and provide marvelous examples of how geometric reasoning continues to light the paths that lead to knowledge previously thought impossible.



Rubik's Cube consists of smaller cubes, where only the outer faces of the outermost cubes can be seen. STEFANO BIANCHETTI/CORBIS.

RUBIK'S CUBE

For any geometric concept, there is an associated puzzle or riddle. While many puzzles have been designed to clearly illustrate a preexisting concept, some sound geometric theories were actually discovered as a solution to such puzzles, or proofs that no solution exists.

Rubik's Cube is possibly the most famous and addictive of all geometric puzzles. It was invented in 1974 by Hungarian sculptor, architectural engineer, and professor, Erno Rubik. Rubik's Cube consists of smaller cubes, where only the outer faces of the outermost cubes can be seen. The original Rubik's Cube has dimensions of $3 \times 3 \times 3$. That is, each edge of the cube is three cube lengths long; each layer of the cube is 3×3 to create a square; and the cube consists of three of these layers; so the measurement in any direction is the length of three smaller cubes. Any layer of the cube contains nine cubes ($3 \times 3 = 3^2 = 9$). There are three layers for a total of 27 smaller cubes ($3 \times 3 \times 3 = 3^3 = 27$).

The Rubik's Cube is a perfect illustration of the reason that numbers like 1, 8, 27, 64, and 125 are referred to as cubic numbers; they can be configured to make perfect cubes. Any eight identical objects can be situated in space to form a cube. This is the geometric interpretation of two raised to the third power. Similarly, if a number of objects can be arranged (in two dimensions) to form a square, that number is called a square number. Square numbers can be represented algebraically by some other whole number raised to the power of two. This concept is illustrated by the faces of a Rubik's Cube.

Professor Rubik was not actually trying to create a puzzle, but rather to solve a three-dimensional geometric problem that had become a hot topic at the time. The problem was to create a seemingly solid cube consisting of twenty-seven smaller cubes, where any layer could be rotated around its center without disturbing the other layers.

Fascinating mechanics allow any layer to be rotated around its center without causing the rest of the apparatus to fall apart. The 27 smaller cubes can be categorized as a single central cube (which is not actually a cube, but the main component of the complex rotating mechanism); six cubes surrounding the central mechanism; and the 20 cubes with faces that can be seen. The rotation of layers in different directions is enabled by a series of spring-loaded spindles and plastic flanges, in addition to the intricate mechanism in the center of the cube.

A Rubik's Cube provides a concrete example of the geometric concepts of surface area and volume. The area of one face on a small cube is equal to the length of one of its edges squared; and each of the six faces of a $3 \times 3 \times 3$ cube consists of nine smaller faces for a total of 54 visible faces; so the surface area of the entire cube is equal to the area of one small face multiplied by 54. The volume of a small cube is equal to the length of one if its

edges cubed; and there are 27 smaller cubes; so the volume of the main cube is equal to the volume of one small cube multiplied by 27. The multitude of mathematical facts that can be illustrated (and even discovered) while playing with a Rubik's Cube is amazing.

Initially and when in solved form, each of the six faces of the cube is its own color: green, blue, red, orange, yellow, or white. As the layers are rotated, the colored faces are shuffled. The goal of the puzzle is to restore each face to a single color after thorough shuffling. Numerous strategies have been developed for solving a Rubik's Cube, all of which involve some degree of geometric reasoning. Some strategies can be simulated by computer programs, and many contests take place to compare strategies based on the average number of moves required to solve randomized configurations. The top strategies can require less than 20 moves.

Possibly the most daunting fact about the $3 \times 3 \times 3$ Rubik's Cube is that 43,252,003,274,489,856,000 different combinations of colors can be created on the faces of the cube. That's more than 43 quintillion combinations, or 43 million multiplied by a million, and then multiplied by a million again. Keep in mind that the original $3 \times 3 \times 3$ cube is among the smallest and least complicated of Rubik's puzzles!

SHOOTING AN ARROW

The aim of archery is to shoot an arrow and hit a target. The three main components involved in shooting an arrow—the bow, the arrow, and the target—are thoroughly analyzed in order to optimize accuracy.

The act of shooting an arrow provides an excellent exploration of vectors (as may be deduced by the fact that vectors are usually represented by arrows in mathematical figures). The intended path of the arrow, the forces that alter this path, and the true path taken by the arrow when released can all be represented as vectors. In fact, the vector that represents the true path taken by the arrow is the sum of the vectors produced by the forward motion of the arrow and the vectors that represent the forces that disrupt the motion of the arrow. Gravity, wind, and rain essentially add vectors to the vector of the intended path, so that the original speed and direction of the arrow is not maintained. When an arrow is aimed directly at a target and then released, it begins to travel in the direction of the target with a specific speed. However, the point at which an arrow is directly aimed is never the exact point hit by the arrow. Gravity immediately adds a downward force to the forward force created by the bow, pulling the arrow down and reducing its speed. Gravity is constant, so the vector used to represent this force always points straight toward the ground with the same magnitude (length). If gravity is the only force acting on an arrow flying toward its target, then the point hit will be directly below the pointed at which the arrow is aimed; how far below depends on the distance the arrow flies. Any amount of wind or rain moving in any direction has a similar affect on the flight of the arrow, further altering the speed and direction of the arrow. To determine the point that the arrow will actually hit involves moving from the intended target in the direction and length of the vectors that represent the additional forces, similar to the way that addition of vectors is represented on a piece of graph paper.

Though the addition of vectors in three-dimensional space is the most prominent application of geometry found in archery, geometric concepts can be unearthed in all aspects of the sport. The bow consists of a flexible strip of material (e.g., wood or light, pliable metal) held at a precise curvature by a taught cord. The intended target and the actual final location of the arrowhead—whether on a piece of wood, a bail of hay, or the ground—can be thought of as theoretical points in space. The most popular target is made of circles with different radial distances from the same center, called concentric circles. If feathers are not attached at precise angles and positions near the rear of the arrow, they will not properly stabilize the arrow and it will wobble unpredictably in flight. In these ways and more, geometric reasoning is essential to every release of an arrow.

STEALTH TECHNOLOGY

Radar involves sending out radio waves and waiting a brief moment to detect the angles from which waves are reflected back. An omnidirectional radar station on the ground detects anything within a certain distance above the surface of Earth, essentially creating a hemisphere of detection range. A radar station in the air (e.g., attached to a spy plane), can send out signals in all directions, detecting any object within the spherical boundary of the radar's range. The direction and speed of an object in motion can be determined by changes in the reflected radio waves. Among other things, radar is used to detect the speed of cars and baseballs, track weather patterns, and detect passing aircraft.

Most airplanes consist almost entirely of round surfaces that help to make them aerodynamic. For example, a cross-section of the main cabin of a passenger plane (parallel to the wingspan or a row of seats) is somewhat circular; so when the plane flies relatively near a radar station on the ground, it provides a perfect reflecting surface for radio waves at all times. To illustrate this, consider

someone holding a clean aluminum can parallel to the ground on a sunny day. If he looks at the can, he will be able to see the reflection of the Sun no matter how the can is turned or moved, as long as it remains parallel to the ground. However, if the can were traded for a flat mirror, he would have to turn the mirror to the proper angle or move it to the correct position relative to his eyes in order to reflect the Sun into his face. The difficulty of accurately reflecting the sun using the flat mirror provides the basis for stealth technology.

To avoid being detected by radar while sneaking around enemy territories, the United States military has developed aircraft—including the B-2 Bomber and the F-117 Nighthawk—that are specially designed to reflect radio waves at angles other than directly back to the source. The underside of an aircraft designed for stealth is essentially a large flat surface; and sharp transitions between the various parts of the aircraft create well-defined angles. The danger of being detected by radar comes into play only if the aircraft is directly above a radar station; a mistake easily avoided with the aid of devices that warn pilots and navigators of oncoming radio waves.

Potential Applications

ROBOTIC SURGERY

While the idea of a robot operating on a human body with metallic arms wielding powerful clamps, prodding rods, probing cameras, razor-sharp scalpels, and spinning saws could make even the bravest of patients squeamish, the day that thinking machines perform vital operations on people may not be that far away.

Multiple robotic surgical aids are already in development. One model is already in use in the United States and another, currently in use in Europe, is waiting to be approved by the U.S. Food and Drug Administration (FDA). All existing models require human input and control. Initial instructions are input via a computer workstation using the usual computer equipment, including a screen and keyboard. A control center is also attached to the computer and includes a special three-dimensional viewing device and two elaborate joysticks. Cameras on the ends of some of the robotic arms near or inside the patient's body send information back to the computer system, which maps the visual information into mathematical data. This data is used to recreate the three-dimensional environment being invaded by the robotic arms by converting the information into highly accurate geometric representations. The viewing device has two goggle-like eyeholes so that the surgeon's eyes and brain perceive the images in three dimensions as well. The images can be precisely magnified, shifting the perception of the surgeon to the ideal viewpoint.

Once engrossed in this three-dimensional representation, the surgeon uses the joysticks to control the various robotic appendages. Pressing a button or causing any slight movement in the joysticks sends signals to the computer, which translates this information into data that causes the precise movement of the surgical instruments. These types of robotic systems have already been used to position cameras inside of patients, as well as perform gallbladder and gastrointestinal surgeries. Immediate goals include operating on a beating heart without creating large openings in the chest.

By programming robotic units with geometric knowledge, humans can accurately navigate just about any environment, from the inside of a beating human heart to the darkest depths of the sea. By combining spacecraft, telescopes, and robotics, scientists can send out robot aids that explore the reaches of the Universe while receiving instructions from Earth. When artificial intelligence becomes a practical reality, scientists in all fields will be able to send out unmonitored helpers to explore any environment, perform tasks, and report back with pertinent information. With the rise of artificial intelligence, robots might soon be programmed to detect any issues inside of a living body, and perform the appropriate operations to restore the body to a healthy state without any human guidance. From the first incision to the final suture, critical decisions will be made by a thinking robotic surgeon.

THE FOURTH DIMENSION

Basic studies in geometry usually examine only three dimensions in order to facilitate the investigation of the properties of physical objects. To say that anything in the Universe exists only in three dimensions, however, is a great oversimplification. As humans perceive things, the Universe has a fourth dimension that can be studied in the same way as the length, width, and height of an object. This fourth dimension is time, and has just as much influence on the state of an object as its physical dimensions. Similar to the way that a cylinder can be seen as a two-dimensional circle extended into a third dimension, a can of soda thrown from one person to another can be seen as a three-dimensional object extending through time, having a different distinct position relative to the things around it at every instant. This is the fundamental concept behind the movement of objects. If there were truly only three dimensions, things could not move

or change. But just as a circular cross-section of a cylinder helps to shed light on its three-dimensional properties, studying snapshots of objects in time makes it possible to understand their structure.

As perceived by the people of Earth, time moves at a constant rate in one direction. The opposite direction in time, involving the moments of the past, only exists in the forms of memory, photography, and scientific theory. Altering the perceived rate of time—in the opposite direction or in the same direction at an accelerated speed—has been a popular fantasy in science fiction for hundreds of years. Until the twentieth century, the potential of time travel was considered by even the most brilliant scientists to lie much more in the realm of fiction. In the last hundred years, however, a string of scientists have delved into this fascinating topic to explore methods for manipulating time.

The idea of time as a malleable (changeable) dimension was initiated by the theory of special relativity proposed by Albert Einstein (1879–1955) in the early twentieth century.

An important result of the theory of special relativity is that when things move relative to each other, one will perceive the other as shrinking in the direction of relative motion. For example, if a car were to drive past the woman in the chair, its length would appear to shrink, but not its height or width. Only the dimension measured in the direction of motion is affected. Of course, humans never actually see this happen because we do not see things that move quickly enough to cause a visible shrinking in appearance. Something would have to fly past the woman at about 80% the speed of light for her to notice the shrinking, in which case she would probably miss the car altogether, and would surely have no perception of its dimensions.

Similar to the manner in which the length of an object moving near the speed of light would seem to shrink as perceived by a relatively still human, time would theoretically seem to slow down as well. However, time would not be affected in any way from the point of view of the moving object, just as physical measurements only seem to shrink from the point of view of someone not moving at the same speed along the same path. If two people are flying by each other in space, to both of these people it will seem that the other is the one moving. So while one could theoretically see physical shrinking and a slowing of the watch on the other's arm, the other sees the same affects in the other person. Without a large nearby reference point, it is easy to feel like the center of the universe, with the movement, mass, and rate of time all-dependent upon the local perception.

All of these ideas about skewed perception due to speed of relative motion are rather difficult to grasp because none of it can be witnessed with human eyes, but recall that the notion of Earth as a sphere moving in space was once commonly tossed aside as mystical nonsense. Einstein's theory of relativity explains events in the Universe much more accurately than previous theories. For example, relativity corrects the inaccuracies of English mathematician Isaac Newton's (1642–1727) proposed laws of gravity and motion, which had been the most acceptable method for explaining the forces of Earth's gravity for hundreds of years. Just as humans can now film the Earth from space to visually verify its spherical nature, its path around the sun, and so forth, the future may very well bring technology that can vividly verify the theories that have been evolving over the last century. For now, these theories are supported by a number of experiments. In 1972, for example, two precise atomic clocks were synchronized, one placed on a high-speed airplane, and the other left on the ground. After the airplane flew around and landed, the time indicated by the clock on the airplane was behind that of the clock on the ground. The amount of time was accurately explained and predicted by the theory of relativity. Inconsistencies in experiments involving the speed of light dating back to the early eighteenth century can be accurately accounted for by the theory of relativity as well.

To travel into the past would require moving faster than the speed of light. Imagine sitting on a spacecraft in outer space and looking through a telescope at someone walking on the surface of Earth. New light is continually reflecting off of Earth and the walker, entering the telescope. However, if the spacecraft were to begin moving away from Earth at the speed of light, the walker would appear to freeze because the spacecraft and the light would be moving at the same speed. The same vision would be following the telescope and no new information from Earth would reach it. The light waves that had passed the spacecraft just before it started moving would be traveling at the same speed directly in front of the spacecraft. If the spacecraft could speed up just a little, it would move in front of the light of the past, and the viewer would again see events from the past. The walker would appear to be moving backward as the spacecraft continued to move past the light from further in the past. The faster the spacecraft moved away from Earth, the faster everything would rewind in front of the viewer's eyes. Moving much faster than the speed of light in a large looping path that returned to Earth could land the viewer on a planet full of dinosaurs. Unfortunately, moving faster than the speed of light is considered to be impossible, so traveling backward in time is out of the

Key Terms

Angle: A geometric figure formed by two lines diverging from a common point or two planes diverging from a common line often measured in degrees.

Area: The measurement of a surface bounded by a set of curves as measured in square units.

Cross-section: The two-dimensional figure outlined by slicing a three-dimensional object.

Curve: A curved or straight geometric element generated by a moving point that has extension only along the one-dimensional path of the point.

Geometry: A fundamental branch of mathematics that deals with the measurement, properties, and relationships of points, lines, angles, surfaces, and solids.

Line: A straight geometric element generated by a moving point that has extension only along the one-dimensional path of the point.

Point: A geometric element defined only by an ordered set of coordinates.

Segment: A portion truncated from a geometric figure by one or more points, lines, or planes; the finite part of a line bounded by two points in the line.

Vector: A quantity consisting of magnitude and direction, usually represented by an arrow whose length represents the magnitude and whose orientation in space represents the direction.

Volume: The amount of space occupied by a three-dimensional object as measured in cubic units.

question. The idea of traveling into the future at and accelerated rate, on the other hand, is believed to be theoretically possible; but the best ideas so far involve flying into theoretical objects in space, such as black holes, which would most likely crush anything that entered and might not even exist at all.

The interwoven relationship of space and time is often referred to as the space-time continuum. To those who possess a firm understanding of the sophisticated ideas of special relativity, the four dimensions of the universe begin to reveal themselves more plainly; and to some, the fabric of time is begging to be ripped in order to allow travel to other times. While time travel is not likely to be realized in the near future, every experiment and theory helps the human race explain the events of the past, and predict the events of the future.

Where to Learn More

Books

Hawking, Stephen. *A Brief History of Time: From the Big Bang to Black Holes*. New York: Bantam, 1998.

Pritchard, Chris. *The Changing Shape of Geometry*. Cambridge, UK: Cambridge University Press, 2003.

Stewart, Ian. *Concepts of Modern Mathematics*. Dover Publications, 1995.

Web sites

Utah State University. "National Library of Virtual Manipulatives for Interactive Mathematics." National Science Foundation. April 26, 2005. <http://matti.usu.edu/nlvm/nav/topic_t_3.html> (May 3, 2005).

Graphing

Overview

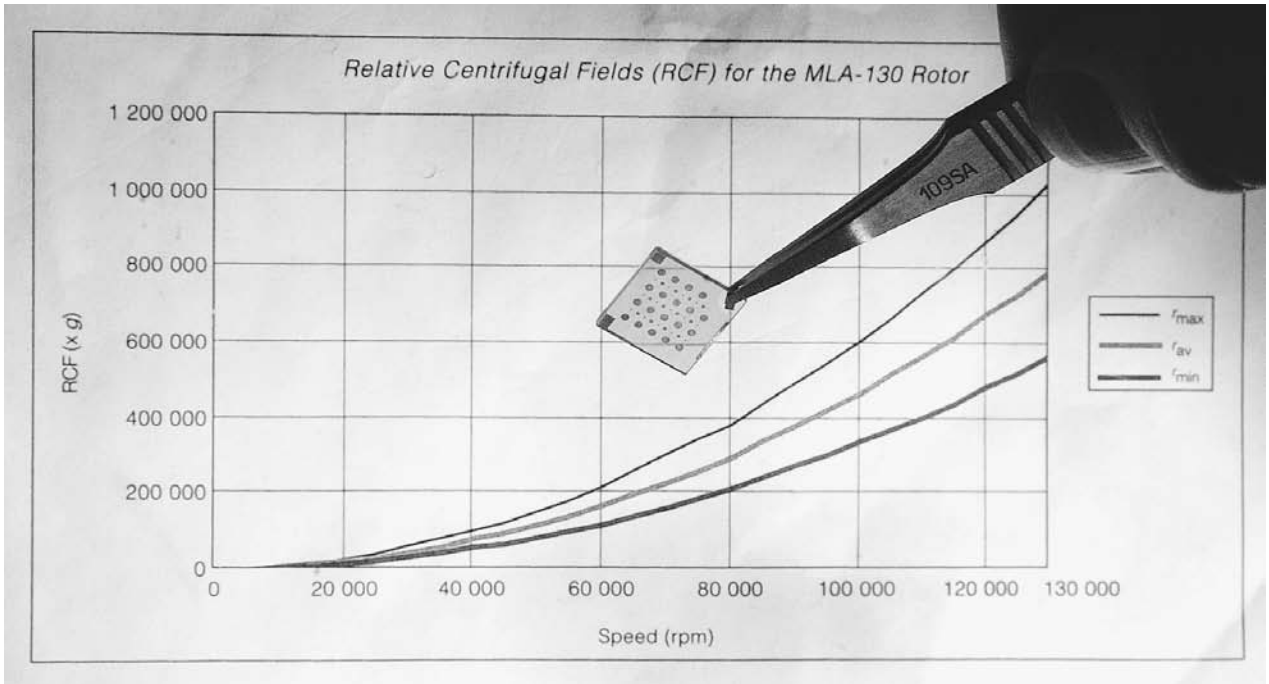
In its most straightforward definition, graphing is the act of representing mathematical relationships or quantities in a visual form. Real-life applications can range from records of stock prices to calculations used in the design of spacecraft to evaluations of global climate change.

Fundamental Mathematical Concepts and Terms

In basic mathematics, graphs depict how one variable changes with respect to another and are often referred to as charts or plots. The graphs can be either empirical, meaning that they show measured or observed quantities, or they can be functional. Examples of empirical measurements are the speed shown on the speedometer of a car, the weight of a person shown on a bathroom scale, or any other value obtained by measurement. Function plots, in contrast, show pure mathematical relationships known as functions, such as $y = b + m, x$, or $y = x^2$. In these examples, each value of x corresponds to a specific value of y and y is said to be a function of x.

Mathematicians and computer scientists sometimes refer to graphs in a different sense when they are analyzing possible ways to connect points (also known as vertices or nodes) in space using networks of lines (also known as edges or arcs). The body of knowledge related to this kind of analysis is known as graph theory. Graph theory has applications to the design of many kinds of networks. Examples include the structure of the electronic links that comprise the Internet, determining the most economical route between two points connected by a complicated network of roads (or railroads, air routes, or shipping routes), electrical circuit design, and job scheduling.

In order to accurately represent empirical or functional relationships between variables, graphs must use some method to scale, or size, the information being plotted. The most common way to do this relies upon an idea developed by the French mathematician René Descartes (1596–1650) in the seventeenth century. Descartes created graphs by measuring the value of one variable along an imaginary line and the value of the second variable along another imaginary line perpendicular to the first. Each of the lines is known as an axis, and it has become standard practice to draw and label the axes rather than using only imaginary lines. Other kinds of coordinate systems exist and are useful for special applications in science and engineering, but the



A computer chip (which contains billions of pure light converting proteins) is shown in the foreground. The chip may one day be a power source in electronics such as mobile phones or laptops. In the background is a graph which displays gravity forces that can separate light-electricity converting protein from spinach. Researchers at MIT say they have used spinach to harness a plant's ability to convert sunlight into energy for the first time, creating a device that may one day power laptops, mobile phones and more. AP/WIDE WORLD PHOTOS. REPRODUCED BY PERMISSION.

majority of graphs encountered on a daily basis use a set of two perpendicular axes.

In most graphs, the dependent variable is plotted using the vertical axis and the independent variable is plotted using the horizontal axis. For example, a graph showing measured rainfall on each day of the year would commonly show the rainfall on the vertical axis because it is dependent upon the day of the year and is, therefore, the dependent variable. Time, represented by the day of the year, is the independent variable because its value is not controlled by the amount of rainfall. Likewise, a graph showing the number of cars sold in the United States for each of the past ten years will usually have the years shown along the horizontal axis and the number of cars sold along the vertical axis. There are some exceptions to this general rule. Atmospheric scientists measuring the amount of air pollution at different altitudes or geologists measuring the chemical composition of rocks at different depths beneath Earth's surface often choose to create graphs in which the independent variable (in these cases, altitude or depth) is shown on the vertical axis. In both cases the dependent variable is being measured vertically, so it makes sense to make graphs having the same orientation.

BAR GRAPHS

Bar graphs are used to show values associated with clearly defined categories. For example, the number of cars sold by a dealer each month, the numbers of homes sold in different cities during a certain year, or the amount of rainfall measured each day during a one-year period can all be shown on bar graphs. The categories are shown along one axis and the values are represented by bars drawn perpendicular to the category axis. In some cases bar graphs will contain a value axis, but in other cases the value axis may be omitted and the values indicated by a number just above or next to each bar. The term "bar graph" is sometimes restricted to graphs in which the bars are horizontal. In that case, graphs with vertical bars are called column graphs.

One bar is drawn for each category on a bar graph, and the height or length of the bar is proportional to the value being shown. For example, the following set of numbers could reflect the average price of homes sold in different parts of Santa Barbara County, California, in February 2005: Area 1, $334,000; Area 2, $381,000; Area 3, $308,000; Area 4, $234,000; Area 5, $259,950. If these figures were plotted on a bar graph, the tallest bar would correspond to the price for Area 2. The absolute height of this

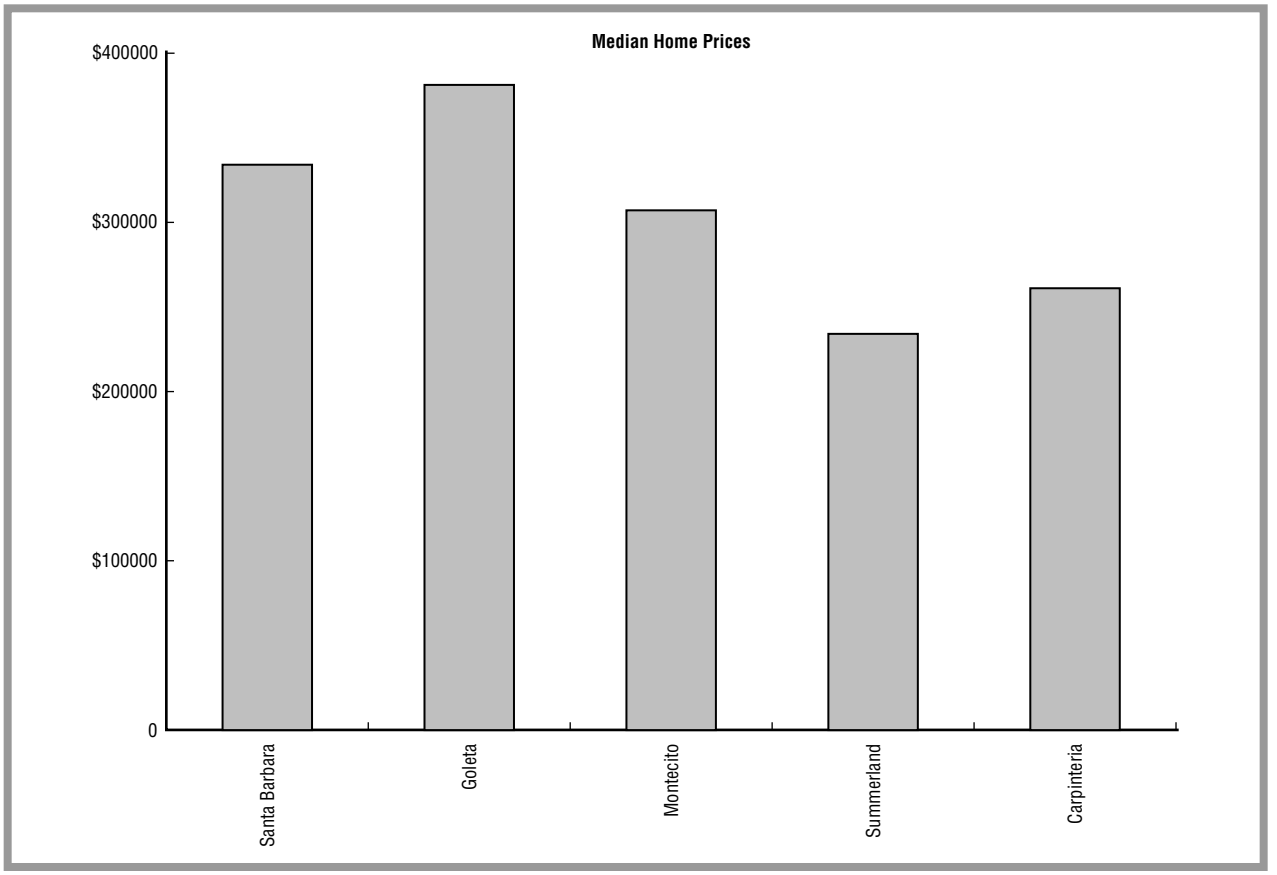


Figure 1.

bar does not matter, because the largest value will control the values of all the other bars. The height of the bar for Area 1, which has the second most expensive homes, would be 334,000 / 381,000 = 88% as tall as the bar representing Area 2. Similarly, the bar representing Area 3 would be 308,000 / 381,000 = 81% as tall as the Area 2 bar. See Figure 1, which depicts the bar graph reflecting the average price of homes sold in different parts of Santa Barbara County, California, in February 2005.

Bar graph categories can represent virtually anything for or about which data can be collected. In Figure 1, the categories represent different parts of a county for which real estate sales records are kept. In other cases bar graph categories represent a quantity such as time, such as the rainfall measured in New York City on each day of February 2005, with each bar representing one day.

Scientists and engineers often use specialized forms of bar graphs known as stem graphs, in which the bars are replaced by lines. Using lines instead of bars can help to make the graph more readable when there are many categories; for example, the sizes of the largest floods along the Rio Grande during the past 100 years would require 100 bars or stems. More often than not, the kinds of data collected by scientists and engineers dictate that the categories involve some measure of distance or time (for example, the year in which each flood occurred). As such, they are usually ordered from smallest to largest. Stem graphs can also have small open or filled circles at the end of each stem. Unless the legend for the graph specifies otherwise, the circles are used simply to make the graph more readable and do not have any significance of their own.

Histograms are specialized bar graphs in which each category represents a range of possible values, and the values plotted perpendicular to the category axis represent the number of occurrences of each category. An important characteristic of a histogram is that each category does not represent just one value or attribute, but rather a range of values that are grouped together into a single category or bin. For example, suppose that in a group of 100 people there are 20 who earn annual salaries between $20,000 and $30,000, 40 who earn annual salaries between $30,001 and $40,000, 30 who earn annual salaries between $40,001 and $50,000, and 10 who earn annual

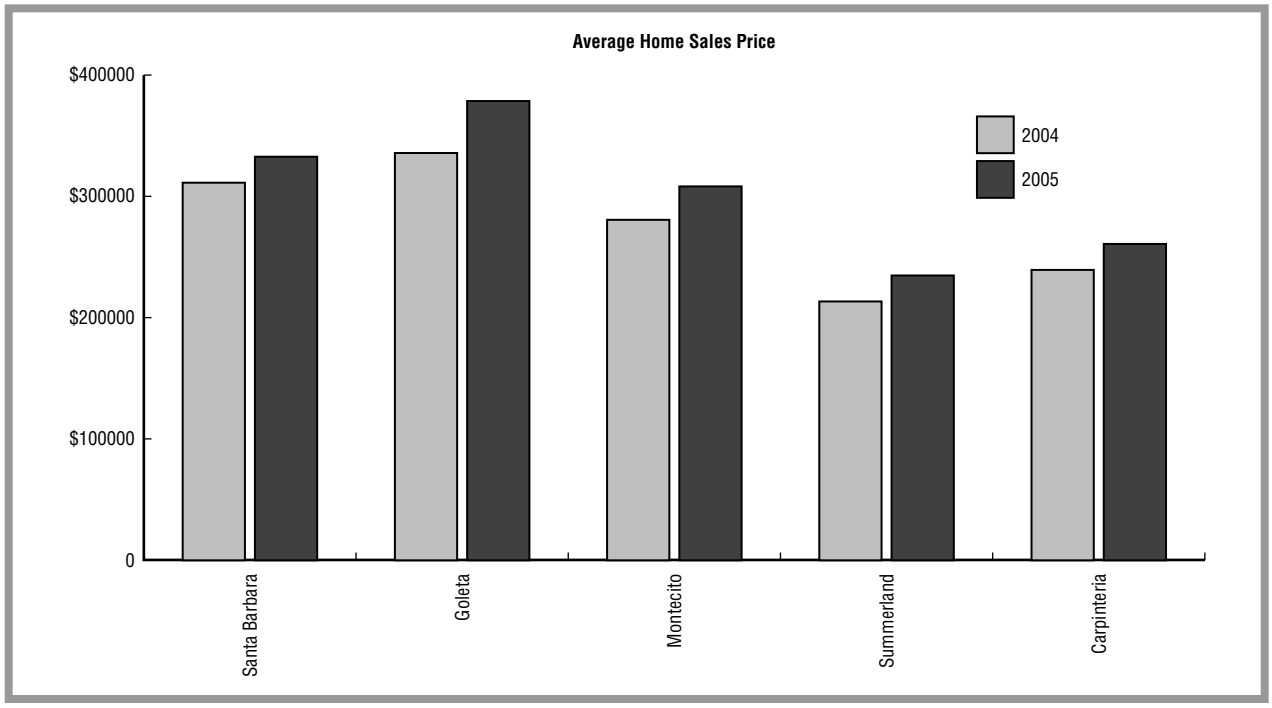


Figure 2.

salaries between $50,001 and $60,000. The bins in a histogram showing this salary distribution would be $20,000 to $30,000, $30,001 to $40,000, $40,001 to $50,000, and $50,001 to $60,000. The height of each bin would be proportional to the number of people whose salaries fall into that bin. The tallest bar would represent the bin with the most occurrences, in this case the $30,001 to $40,000. The second tallest bar would represent the $40,001 to $50,000 category, and it would be 30/40 = 75% as tall as the tallest bin. The width of each bin is proportional to the range of values that it represents. Therefore, if each class interval is the same size then all of the bars on a histogram will be the same width. A histogram containing bars with different widths will have unequal class intervals.

Some bar graphs use more than one set of bars in order to convey several sets of information. Continuing with the home price example from Figure 1, the bars showing the 2005 prices could be supplemented with bars showing the average home sales prices for the same areas in February 2004. Figure 2 allows readers to quickly compare prices and see how they changed between 2004 and 2005. Each category has two bars, one for 2004 and one for 2005, filled with different colors, patterns, or shades of gray to distinguish them from each other.

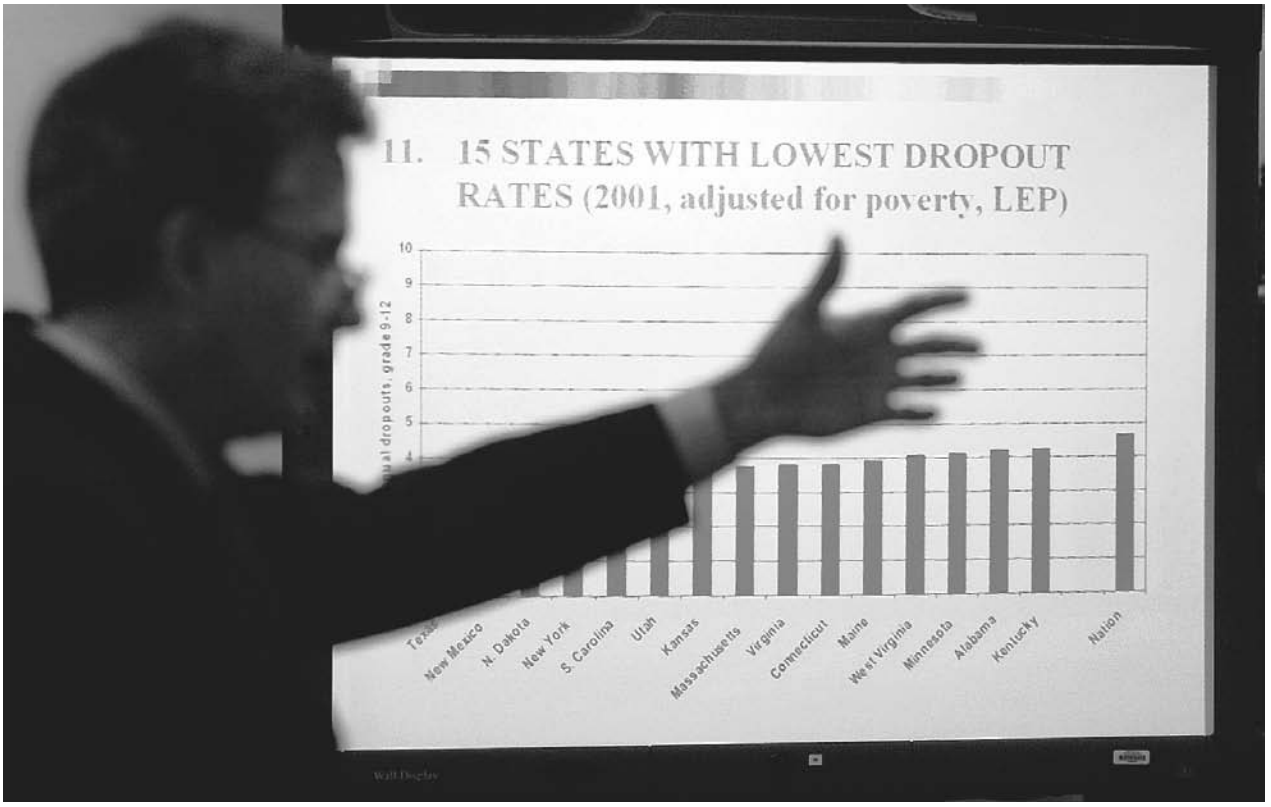
A third kind of bar graph is the stacked bar graph, in which different types of data for each category are represented using bars stacked on top of each other. The bottom bar in each of the stacks will generally have a different height, which makes it difficult to compare values among categories for all but the bottom bars. For this reason, stacked bar graphs can be difficult to read and should generally be avoided.

LINE GRAPHS

Line graphs share some similarities with bar graphs, but use points connected by straight lines rather than bars to represent the values being graphed. As with bar graphs, the categories on a line graph can represent either some kind of measurable quantity or more abstract qualities such as geographic regions.

Line graphs are constructed much like bar graphs. In line graphs, values for each category are known or measured, and the categories are placed along one axis. The values are then scaled along the value axis, and a point, sometimes represented by a symbol such as a circle or a square, is drawn to represent the value for each category. The points are then connected with straight line segments to create the line graph.

One of the weaknesses of line graphs is that they can imply some kind of connection between categories, which may or may not be the intention of the person creating the graph. In a bar chart, each category is represented by a bar that is completely separate from its



Graphs are often used as visuals representing finances. AP/WIDE WORLD PHOTOS. REPRODUCED BY PERMISSION.

neighbors. Therefore, no connection or relationship between adjacent categories is implied by the graph. A line graph implies that the value varies continuously between adjacent categories because the points are connected by lines. If there is no real connection between the values for adjacent categories, for example the home sales prices used in the Figure 1 bar graph example, then it may be better to use a bar graph or stem graph than a line graph.

Like bar graphs, line graphs can be combined to create multiple line graphs. Each line represents a different value associated with each category. For example, a multiple line graph might show different household expenses for each month of the year (rent, heat, water, groceries, etc.) or the income and expenses of a business for each quarter of a particular year. Rather than being placed side-by-side as in a multiple bar graph, however, multiple line graphs are placed on top of each other and the lines are distinguished by different colors or patterns. If only two sets of values are being graphed and their values are significantly different, two value axes may be used. As shown in Figure 3, each value axis corresponds to one of the sets of values and is labeled accordingly.

AREA GRAPHS

Area graphs are line graphs in which the area between the line and the category axis is filled with a color or pattern, and are used when there is a need to show both the values associated with each category and the total of all the values. As Figure 4 shows, the values are represented by the height of the colored area, whereas the total is represented by the amount of area that is colored. If the total area beneath the lines is not important, then a bar graph or line graph may be a better choice. Area graphs can also be stacked if the objective is to show information about more than one set of values. The result is much like a stacked bar graph.

PIE GRAPHS

Pie graphs are circular graphs that represent the relative magnitudes of different categories of data using angular wedges resembling slices of pie. The size of each wedge, which is measured as an angle, is proportional to the relative size of the value it represents.

If the data are given as percentages that add up to 100%, then the angular increment of each wedge is its

percentage × 360°, which is the number of degrees in a complete circle. For example, imagine that Store A sells 30% of all computers sold in Boise, Idaho, Store B sells 18%, and all other stores combined sell the remainder. The wedge representing Store A would be 0.30 × 360° = 108° in size. The wedge representing Store B would, by the same logic, be 0.18 × 360° = 65°, and the wedge representing all other stores would (1.00 − 0.30 − 0.18) × 360° = 0.52 × 360° = 187°. Figure 5 depicts a representative pie graph.

The calculations become slightly more complicated if the data are not given in terms of percentages that add up to 100%. Suppose that instead of the percentage of computers sold by the stores in the previous example, only the number of computers sold by each store is known. In that case, the number of computers sold by each store must be divided by the total number sold by all stores to calculate the percentage for that store. If Store A sold 1,500 computers, Store B sold 900 computers, and all other stores combined sold 2,600 computers, then the total number of computers sold would be 5,000. The percentage sold by Store A would be 1,500/5,000 = 0.30, or 30%. Similar calculations produce results of 18% for Store B and 52% for all other stores combined (just as in the previous example).

RADAR GRAPHS

Radar graphs, also known as spider graphs or star graphs, are special types of line graphs in which the values are plotted along axes radiating from a common point. The result is a graph that looks like a radar screen to some people, and a spider or star to others. There is one axis for each category being graphed, so for n categories each axis will be separated by an angle of $360°/n$. A radar graph showing five categories, for example, would have five axes separated by angles of 360°/5 = 72°. The value of each category is measured along its axis, with the distances from the center proportional to the value, and adjacent values are connected to form an irregularly shaped polygon. One of the advantages of radar plots, as shown below in Figure 6 (p. 254), is that they can convey information about the values of many categories using shapes (the polygons created by connecting adjacent values) that can be easily compared for many different data sets.

Multiple radar graphs are constructed much like multiple line graphs, with several values plotted for each category. The lines connecting the values for each category have different colors or patterns in order to distinguish among them.

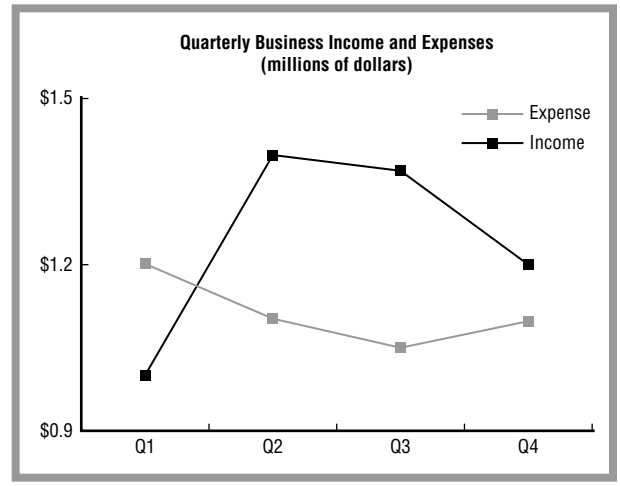


Figure 3.

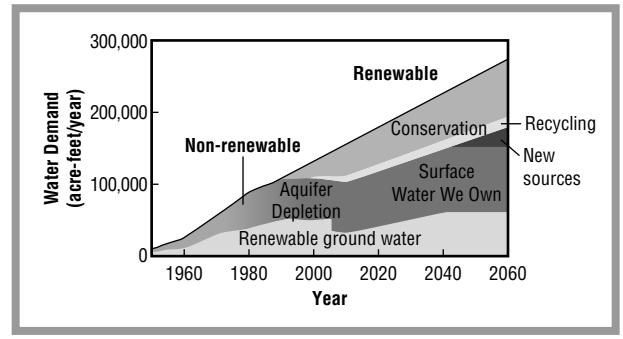


Figure 4: Stacked area graph showing different sources of water (values) by year (categories).

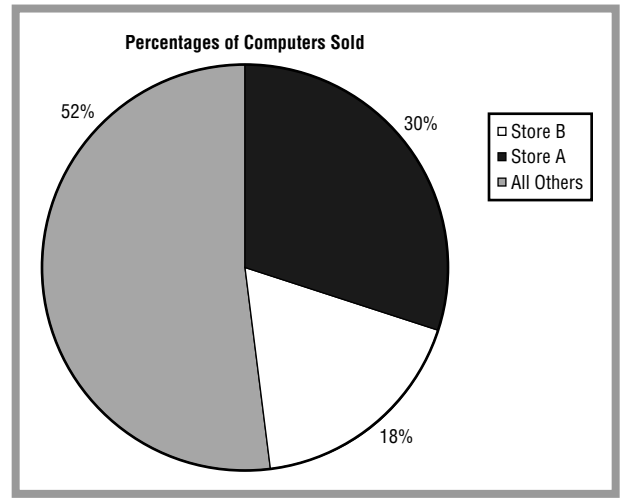


Figure 5.

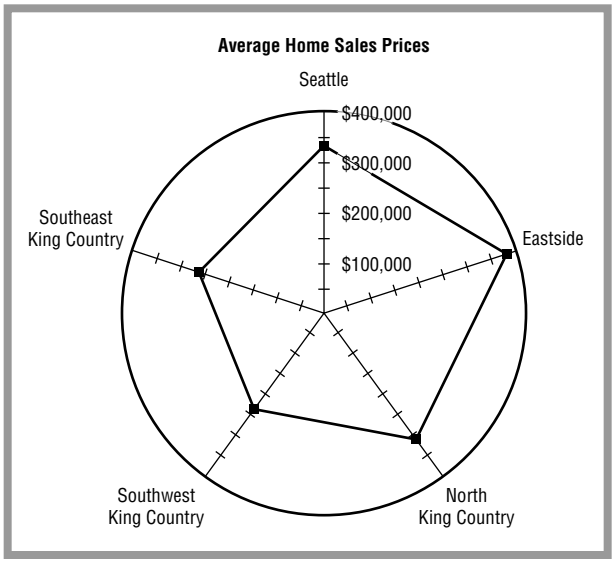


Figure 6.

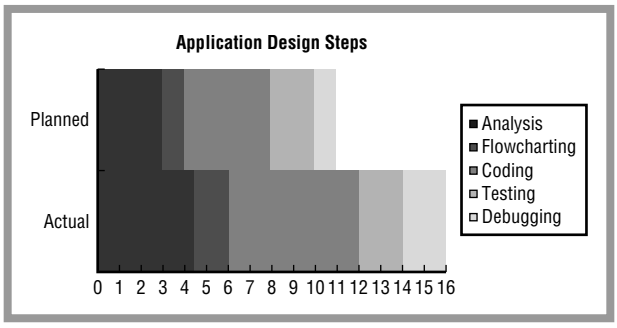


Figure 7.

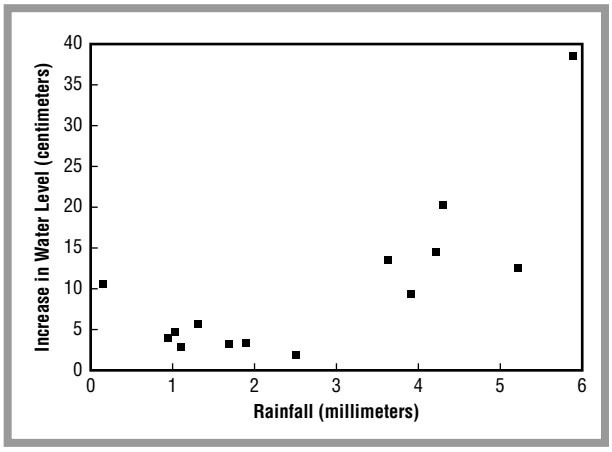


Figure 8.

GANTT GRAPHS

Gantt graphs are used by project managers and others to show job activity over time, which can range from a single workday to a complicated construction project that stretches over several years. The horizontal axis shows time, with units depending on the length of the project. The vertical axis shows resources, which can be anything from the names of people working on the project to different pieces of equipment needed to complete the project. Blocks of time are marked off along the time axis showing how each resource will be used during that time.

PICTURE GRAPHS

Graphs that are intended for general readers rather than scientists or engineers, such as those frequently published in newspapers and magazines, often use artistic symbols to denote the values of different categories. An article about money, for example, might show stacks of currency instead of plain bars in a bar graph. A different article about new car sales might include a graph using a small picture of a car to represent every 100 cars sold by different dealers. These kinds of artistic graphs are usually varieties of bar graphs, although the use of artistic symbols can make it difficult to accurately compare values among different categories. Therefore, they are most useful when used to illustrate general trends or relationships rather than to allow readers to make exact comparisons. For that reason, picture graphs are almost never used by scientists and engineers.

X-Y GRAPHS

X-y graphs are also known as scatterplots. Instead of having a fixed number of categories along one axis, x-y graphs allow an infinite number of points along two perpendicular axes and are used extensively in scientific and engineering applications. Each point is defined by two values: the abscissa, which is measured along the x axis, and the ordinate, which is measured along the y axis. Strictly speaking, the terms abscissa and ordinate refer to the values measured along each axis although in day-to-day conversation many scientists and engineers use the terms in reference to the axes themselves. Each piece of data to be graphed will have both an abscissa and an ordinate, sometimes referred to as x- and y-values.

The most noticeable property of an x-y graph is that it consists of points rather than bars or lines. Lines can be added to x-y plots but they are in addition to the points and not a replacement for them. Line graphs can also have points added as an embellishment and can therefore

Graphing Functions and Inequalities

Continuous mathematical functions and inequalities involving real numbers have an infinite number of possible values, but are graphed in much the same way as x-y graphs containing a finite number of points.

Consider the function $y = x^2$. The first step is to determine the range of the x axis because, unlike a finite set of points that have a minimum and maximum x value, functions can generally range over all possible values of x from $-\infty$ to $+\infty$. For this example, allow x to range from 0 to 3 ($0 \geq x \geq 3$). Next, select enough points over that range to produce a smooth curve. This must be done by trial and error, and becomes easier once a few graphs are made. Seven points will suffice for this example: 0, 0.5, 1, 1.5, 2, 2.5, and 3. These values will be the abscissae. Substitute each abscissa into the function (in this case $y = x^2$) and calculate the value of the function for that value, which will produce the ordinates 0, 0.25, 1, 2.25, 4, 6.25, and 9. Finally, plot a point for each corresponding abscissa and ordinate, or (0,0), (0.5,0.25), (1,1), (1.5,2.25), (2,4), (2.5,6.25), and (3,9).

Because a continuous function has values for all possible values of x, not just those for which values were just calculated, the points can be joined using a smooth curve. Before computers with graphics capabilities were widely available, this was done using drafting templates known as French curves, or thin flexible strips known as splines. The French curve, or spline, was positioned so that it passed through the graphed points and used as a guide to draw a smooth curve. A smooth curve can also be approximated by calculating values for a large number of points and then connecting them with straight lines, as in a line graph. If enough points are used, the straight line segments will be short enough to give the appearance of a smooth curve. Computer graphics programs follow a digital version of this procedure, calculating enough sets of abscissae and ordinates to generate the appearance of a continuous line. In many cases the programs use sophisticated algorithms that minimize the number of points by evaluating the function to see where values change the most, plotting more points in those areas and fewer in parts of the graph where the function is smoother.

To plot an inequality, temporarily consider the inequality sign ($<$, $>$, $\geq$, $\leq$) to be an equal sign. Decide upon a range for the abscissae, divide it into segments, and calculate pairs of abscissae and ordinates in the same manner as for a function. If the inequality is $>$ or $<$, then connect the points with a dashed line and indicate which side of the line represents the inequality. For example, if the inequality is $y > x^2$, then the area above the dashed line should be shaded or otherwise identified as the region satisfying the inequality. If the inequality had been $y < x$, then the area beneath the dashed line would satisfy the inequality. In cases of $\geq$ or $\leq$ inequalities, the two regions can be separated by a solid line to indicate that points exactly along the line, not just those above or below it, satisfy the relationship.

Graphs of functions can also be used to solve equations. The equation $4.3 = x^2$, for example, is a version of the equation $y = x^2$ described in this sidebar. Therefore, it can be solved by graphing the function $y = x^2$ over a range of values that includes $x = 4.3$ (for example, $4 \geq x \geq 5$) and reading the abscissa that corresponds to an ordinate of 4.3. In this case, the answer is $x = 2.07$.

be confused with x-y graphs under some circumstances. Line graphs and x-y graphs, however, have some important differences. First, the categories on a line graph do not have to be numbers. As described above, line graphs can represent things such as cities, geographic areas, or companies. Each value on a line graph must correspond to one of a finite number of categories. The abscissa of a point plotted on an x-y graph, in contrast, must always be a number and can take on any value. Second, the lines on a line graph must always connect the values for each category. If lines are added to an x-y graph, they do not have to connect all of the points. Although they can connect all of the points, especially in cases where there are only a few points on the graph, lines connecting the data points are not required on x-y graphs. Lines can, for example, be used to show averages or trends in the data on an x-y graph. Figure 8 represents an x-y graph. Adding lines to connect all of the points in an x-y graph can be very confusing if there are a large number of points, and should be done only if it improves the legibility of the graph.

To create an x-y graph, first move along the x-axis to the abscissa and draw an imaginary line perpendicular to the x-axis and passing through the abscissa. Next, move

Graphing Fallacies

Some people believe that graphs don't lie because they are based on numbers. But, the way that a graph is drawn and the numbers that are chosen can deliberately or accidentally create false impressions of the relationships shown on the graph. Scientists, engineers, and mathematicians are usually very careful not to mislead their readers with fallacious graphs, but artists working for newspapers and magazines sometimes take liberties that accidentally misrepresent data. Dishonest people may also deliberately create graphs that misrepresent data if it helps them to prove a point.

One way to misrepresent data is to create a graph that shows only a selected portion of the data. This is known as taking data out of context. For example, if the number of computers sold at an electronics store increases by 100 computers per year for four years and then decreases by 25 computers per year during the fifth year, it is possible to make a graph showing only the last year's information and title the graph, "Decreasing Computer Sales." Actually, though, sales have increased by $4 \times 100 - 25 = 375$ computers over the five years, so the fifth year represents only a small change in a longer term trend. It is true to state that computer sales fell during the fifth year but, depending on how the graph is used, it may be misleading to do so because it presents data out of context.

Another way to misrepresent data is by choosing the limits of the vertical axis of the graph. Imagine that a survey shows that men working in executive jobs earned an average salary of $100,000 per year and that women working in executive jobs earned an average salary of $85,000 per year. If these two pieces of information were plotted on a graph with an axis ranging from zero to $100,000, it would be clear that the women earned an average of 15% less than the men. But, if the axis were changed so that it ranged only from $80,000 to $100,000 it might appear to the casual reader than women earned only about 25% as much as men. Because the information conveyed by a graph is largely visual, many readers will not notice the values on the axis and base their interpretation only on the relationships among the lines, bars, or points on the graph. Some irresponsible graph-makers even eliminate the ordinate axis altogether and use bars or other symbols that are not proportional to the values that they represent.

Sometimes it is the data themselves that are the problem. A graph showing how salaries have increased during the past 50 years may show a tremendous increase. If the salaries are adjusted for inflation, however, the increase may appear to be much smaller.

along the y-axis to the ordinate, then draw an imaginary line perpendicular to the y-axis. Draw a small symbol at the location where the two imaginary lines intersect. Repeat this procedure for each of the points to be graphed. The symbols used should be the same for all of the points in each data set, and can be circles, squares, rectangles, or any other simple shape. If more than one data set is to be shown on the same graph, choose a different symbol or color for the points in each set.

The abscissa and ordinate values of points on x-y graphs created for scientific or engineering projects are sometimes transformed. This can be done in order to show a wide range of values on a single set of axes or, in some cases, so that points following a curved trend are graphed as a straight line. The most common way to transform data is to calculate the logarithm of the abscissa or ordinate, or both. If the logarithm of one is plotted against the original arithmetic value of the other, the graph is known as a semi-log graph. If the logarithms of both the abscissa and ordinate are plotted, the result is a log-log graph. The logarithms used can be of any base, although base 10 is the most common, and the base should always be indicated. At one time, base 10 logarithms were referred to as common logarithms and denoted by the abbreviation log. Base e logarithms ($e = 2.7183\ldots$) were referred to as natural logarithms and denoted by the abbreviation ln. This practice fell out of favor among some scientists and engineers during the late 1900s. Since then, it has been common to use log to denote the natural logarithm, and $\log_{10}$ to denote the base 10, or common, logarithm.

A map with points plotted to indicate different cities or landmarks can be considered to be a special kind of x-y graph. In this case, the abscissa and ordinate of each point consist of its geographic location given in terms of latitude and longitude, universal transverse Mercator (UTM) coordinates, or other cartographic coordinate systems. Likewise, the outline of a country or continent can be thought of as a series of many points connected by short line segments.

The underlying principles of x-y plots can be extended into the third dimension to produce x-y-z plots. Points are plotted along the z axis following the same procedure that is used for the x and y axes. One difficulty associated with x-y-z plots is that two-dimensional surfaces such as pieces of paper have only two dimensions. Complicated geometric constructions known as projections must be used to create the illusion of a third dimension on a flat surface. Therefore, x-y-z plots of large numbers of points are practical only if done on a computer, which allows the plots to be virtually rotated in space so that the data can be examined from any perspective.

BUBBLE GRAPHS

Bubble graphs allow three-dimensional data to be presented in two-dimensional graphs, and are in many cases useful alternatives to x-y-z graphs. For each data point, two of the three variables are plotted as in a normal x-y graph. The third variable for each point is represented by changing the size of the point to create circles or bubbles of different sizes. One important consideration is the way in which the bubble size is calculated. One way is to make the diameter of the circle proportional to the value of the third variable. Because the area of a circle is proportional to the square of its radius, doubling the radius or diameter will increase the area of the circle by a factor of 4. Therefore, doubling the diameter may mislead a reader into believing that one bubble represents a value four times as large as another when the person creating the graph intended it to represent a value only twice as large. In order to create a circle with twice the area, the radius or diameter must be increased by a factor of 1.414 (which is the square root of 2). Figure 9 is representative of a bubble graph.

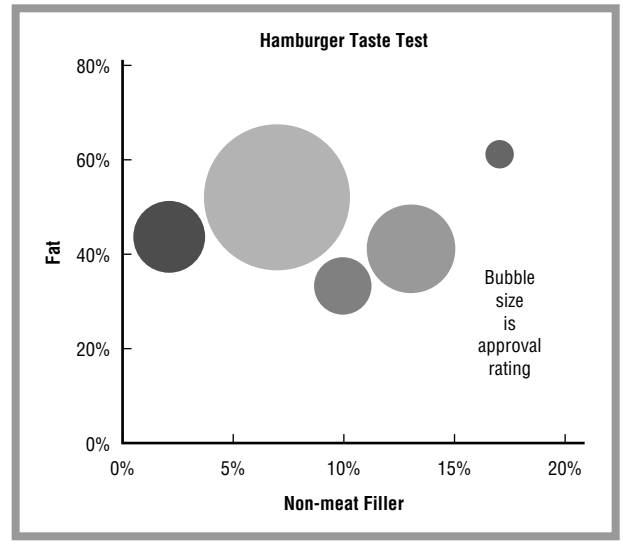


Figure 9.

A Brief History of Discovery and Development

The graphing of functions was invented by the French mathematician and philosopher René Descartes (1596–1650) in 1637, and the Cartesian coordinate system of x-y (and sometimes z) axes used to plot most graphs today bears his name. Ironically, however, Descartes did not use axes as known today or negative numbers when he created the first graphs.

Commercially manufactured graph paper first appeared in about 1900 and was adopted for use in schools as part of a broader reform of mathematics education. Leading educators of the day extolled the virtues of using so-called squared paper or paper with squared lines to graph mathematical functions. As the twentieth century progressed, students and professionals came to have a wide range of specialized graph paper available for use. The selection included graph paper with preprinted semi-log and log-log axes, as well as paper designed for special kinds of statistical graphs.

Digital computers were invented in the middle of the twentieth century, but computers capable of displaying even simple graphs were rare until personal computers became common in the 1980s. So-called spreadsheet programs, in particular, represented a great advance because they allowed virtually anyone to enter rows and columns of numbers and then examine relationships among them by creating different kinds of graphs. Handheld graphing calculators appeared in the 1990s and were quickly incorporated into high school and college mathematics courses. At about the same time, sophisticated scientific graphing and visualization programs for advanced students and professionals began to appear. These programs could plot thousands of points in two or three dimensions.

Real-life Applications

GLOBAL WARMING

Most scientists studying the problem have concluded that burning fossil fuels such as coal and oil (including gasoline) during the twentieth century has caused the amount of carbon dioxide, carbon monoxide, and other gasses in Earth's atmosphere to increase, which has in turn led to a warming of the atmosphere and oceans. Among

the tools that scientists use to draw their conclusions are graphs showing how carbon dioxide and temperature change from day to day, week to week, and year to year. Although actual measurements of atmospheric gasses date back only 50 years or so, paleoclimatologists use other information such as the composition of air bubbles trapped for thousands of years in glacial ice, the kinds of fossils found buried in lake sediments, and the widths of tree rings to infer climate back into the recent geologic past. Data collected over time are often described as time series. Time series can be displayed using line graphs, stem graphs, or scatter plots to illustrate both short-term fluctuations that occur from month to month and long-term fluctuations that occur over tens to thousands of years, and have provided compelling evidence that increases in greenhouse gasses and temperatures measured over the past few decades represent a significant change.

FINDING OIL

Few oil wells resemble the gushers seen in old movies. In fact, modern oil well-drilling operations are designed specifically to avoid gushers because they are dangerous to both people and the environment. Geologists carefully examine small fragments of rock obtained during drilling and, after drilling is completed, lower instruments down the borehole to record different rock properties. These can include electrical resistivity, natural radioactivity, density, and the velocity with which sound waves move through the rock. All of this information helps to determine if there is oil thousands of feet beneath the surface, and is plotted on special graphs known as geophysical logs. In most cases, the properties are measured once every 6 inches (15.2 cm) down the borehole, so depth is the category (or abscissa) and each rock property is a value (or ordinate). Unlike most line graphs or x-y graphs, though, the category axis or abscissa is oriented vertically with the positive end pointing downward because the borehole is vertical and depth is measured from the ground surface downward. Geophysical logs are plotted together on one long sheet of paper or a computer screen so that geologists can compare the graphs, analyze how the rock properties change with depth, and then estimate how much oil or gas there is likely to be in the area where the well was drilled. If there is enough to make a profit, pipes and pumps are installed to bring the oil to the ground surface. If not, the well is called a dry hole and filled with cement.

GPS SURVEYING

Surveyors, engineers, and scientists use sensitive global positioning system (GPS) receivers that can determine the locations of points on Earth's surface to an accuracy of a fraction of an inch. In some cases, the information is used to determine property boundaries or to lay out construction sites. In other cases, it is used to monitor movements of Earth's tectonic plates, the growth of volcanoes, or the movement of large landslides. GPS users, however, must be certain that their receivers can obtain signals from a sufficient number of the 24 global positioning system satellites orbiting Earth in order to make such accurate and precise measurements. This can be difficult because the number of satellites from which signals can be received in a given location varies from place to place throughout the course of the day. Professional GPS users rely on mission-planning software to schedule their work so that it coincides with acceptable satellite availability. Two of the most important pieces of information provided by mission-planning software are bar graphs showing the number of satellites from which signals can be received and the overall quality or strength of the signals, which is known as positional dilution of precision (PDOP). A surveyor or scientist planning to collect high-accuracy GPS measurements will enter the latitude and longitude of the project area, information about obstructions such as tall buildings or cliffs, and the date the work is to take place. The mission-planning software will then create a graph showing the satellite coverage and PDOP during the course of that day, so that fieldwork can be scheduled for the most favorable times.

BIOMEDICAL RESEARCH

Genetic and biomedical research generate large amounts of data, particularly related to genetic sequences or genomes. Researchers in these fields use specialized graphing programs to visualize genetic sequences of different organism, including computer programs that can simultaneously display information about two different organisms and graphically illustrate which genes are present in both. Phylogenetic tree graphs, which have a branching structure, are used to illustrate the relationships between groups of many different organisms. Other biomedical scientists have developed new ways to construct multidimensional graphs to represent similarities between proteins. The field of biomechanics combines physics with biology and medicine to analyze how physical stresses and forces affect living organisms. Sophisticated scientific visualization software is used to analyze computer models simulating the stresses developed in the bones of athletes or in the blood vessels of people suffering heart attacks.

Technical Stock Analysis

Some investors rely on hunches or tips from friends to decide when they should buy or sell stock. Others rely on technical analysis to spot trends in stock prices and sales that they hope will allow them to earn more money by buying or selling stock at just the right time. Technical stock analysts use different kinds of specialized graphs to depict information that is important to them. Candlestick plots use one symbol for each day to show the price of the stock when the market opened, the price when it closed, and the high and low values for the stock during the course of the day. This is done by using a rectangle to indicate opening and closing prices, with vertical lines extending upward and downward from the box to indicate the daily high and low prices. The result is a symbol that looks like a candle with a wick at each end. The color of the box, usually red or green, indicates whether the closing price of the stock was higher or lower than the opening price.

Day-to-day fluctuations in stock price can be smoothed out using moving average or trend plots that remove most, if not all, of the small changes and let investors concentrate on trends that persist for many days, weeks, or even months. Moving averages calculate the price of a stock on any given day by averaging the prices over a period of days. For example, a five-day moving average would calculate the average price of the stock over a five-day period. The "moving" part of moving average means that different sets of data are used to calculate the average each day. The five-day moving average calculated for June 5, 2004, will use a different set of five prices (for June 1 through June 5) than the five-day moving average calculated for June 6, 2004 (June 2 through June 6).

The volume, or number, of shares sold on a given day, is also important to stock analysts and can be shown using bar charts or line graphs.

PHYSICAL FITNESS

Many health clubs and gyms have a variety of computerized machines such as stationary bicycles, rowing machines, and elliptical trainers that rely on graphs to provide information to the person using the equipment. At the beginning of a workout, the user can scroll through a menu of different simulated routes, some hilly and some flat, that offer different levels of physical challenge. As the workout progresses, a bar graph moves across a small screen to show how the resistance of the machine changes to simulate the effect of running or bicycling over hilly terrain. In other modes, the machine might monitor the user's pulse and adjust the resistance to maintain a specified heartbeat, with the level of resistance shown using a different bar graph.

AERODYNAMICS AND HYDRODYNAMICS

The key to building fast and efficient vehicles—whether they are automobiles, aircraft, or watercraft—lies in the reduction of drag. Using a combination of experimental data from wind tunnels or water tanks and the results of computational fluid dynamics computer simulations, designers can create graphs showing how factors such as the shape or smoothness of a vehicle affect the drag exerted by air or water flowing around the vehicle. Experiments are conducted or computer simulations run for different vehicle shapes, and the results are summarized on graphs that allow designers to choose the most efficient design for a particular purpose. In some cases, these are simply x-y graphs or line graphs comparing several data sets. In other cases, the graphs are animated scientific visualizations that allow designers to examine the results of their experiments or models in great detail.

COMPUTER NETWORK DESIGN

Computer networks from the Internet to the computers in a small office can be analyzed using graphs showing the connectivity of different nodes. A large network will have many nodes and sub-nodes that are connected in a complicated manner, partly to provide a degree of redundancy that will allow the network to continue operating even if part of it is damaged. The United States government funded research during the 1960s on the design of networks that would survive attacks or catastrophes grew into the Internet and World Wide Web. A network in which each computer is connected to others by only one pathway, be it a fiber optic cable or a wireless signal, can be inexpensive but prone to disruption. At the other end of the spectrum, a network in which each computer is connected to every other computer is almost

Scientific Visualization

Scientific visualization is a form of graphing that has become increasingly important since the 1980s and 1990s. Advances in computer technology during those years allowed scientists and engineers to develop sophisticated mathematical simulations of processes as diverse as global weather, groundwater flow and contaminant transport beneath Earth's surface, and the response of large buildings to earthquakes or strong winds. Likewise, computers enabled scientists and engineers to collect very large data sets using techniques like laser scanning and computerized tomography. Instead of tens or hundreds of points to plot in a graph, scientists working in 2005 can easily have thousands or even millions of data points to plot and analyze.

Scientific visualizations, which can be thought of as complicated graphs, usually contain several different data sets. A visualization showing the results from a computer simulation of an oil reservoir, for example, might include information about the shape and extent of the rock layers in which the oil is found, information about the amount of oil at different locations in the reservoir, and information about the amount of oil pumped from different wells. A visualization of a spacecraft reentering Earth's atmosphere might include the shape of the spacecraft, colors to indicate the temperature of the outside of the spacecraft, and vectors or streamlines showing the flow of air around the spacecraft. Animation can also be an important aspect of scientific visualization, especially for problems in which the values of variables change over time. Visualization software available in 2005 typically allows scientists to interactively rotate and zoom in and out of plots showing several different kinds of data in three dimensions.

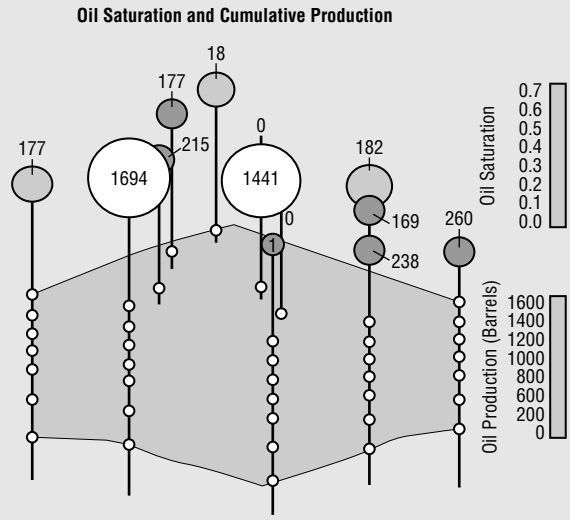


Figure A: Scientific visualization, especially for problems in which the values of variables change over time such as representations of data related to oil drilling depicted above, are an increasingly important ways to understand and depict data.

always prohibitively expensive even though it may be the most reliable. Therefore, the design of effective networks balances the costs and benefits of different alternatives (including the consequences of failure) in order to arrive an optimal design. Because of their built-in redundancy and complexity, large computer networks are impossible to comprehend without graphs illustrating the degrees of interconnection between different nodes. Applied mathematicians also use graph theory to help design the most efficient networks possible under a given set of constraints.

Potential Applications

The basic methods of graphing have not changed over the years, but continually increasing computer capabilities give scientists, engineers, and businesspeople powerful and flexible graphing tools to visualize and analyze large amounts of data. Likewise, scientific visualization tools provide a way to comprehend the voluminous output of supercomputer models of weather, ocean circulation, earthquake activity, climate change, and other complicated natural processes. Ongoing technology development is concentrated on the use of larger and faster computers to better visualize these kinds of data sets, for example using transparent surfaces and advanced rendering techniques to visualize three-dimensional data. Computer-generated movies or animations will also allow visualization of changes in three-dimensional data sets over time (so-called four-dimensional analysis). The design and implementation of user-friendly interfaces will also continue, bring powerful visualization technology within the grasp of more people.

Where to Learn More

Books

Few, Stephen. *Show Me the Numbers: Designing Tables and Graphs to Enlighten.* Oakland, CA: Analytics Press, 2004.

Huff, Darrell. *How to Lie with Statistics.* New York: W.W. Norton, 1954.

Tufte, E.R. *The Visual Display of Quantitative Information.* Cheshire, CT: Graphics Press, 1992.

Web sites

Friendly, Michael. "The Best and Worst of Statistical Graphics." Gallery of Data Visualization. 2000. <http://www.math.yorku.ca/SCS/Gallery/> (March 9, 2005).

Goodman, Jeff. "Math and Media: Deconstructing Graphs and Numbers." How Numbers Tell a Story. 2004. <http://www.ced.appstate.edu/~goodmanj/workshops/ABS04/graphs/graphs.html> (March 9, 2005).

National Oceanic and Atmospheric Administration. "Figures." Climate Modeling and Diagnostics Laboratory. <http://www.cmdl.noaa.gov/gallery/cmdl_figures> (March 9, 2005).

Weisstein, E.W. "Function Graph." Mathworld. <http://mathworld.wolfram.com/FunctionGraph.html> (March 9, 2005).

Imaging

Overview

We each process hundreds or thousands of manufactured images every day, including those displayed by books, magazines, computers, digital cameras, signage, TVs, and movies. Images are an important form of communication in entertainment, war, science, art, and other fields because a human being can grasp more information more quickly by looking at an image than in any other way.

Fundamental Mathematical Concepts and Terms

Most of the images we see have been either altered or created from scratch using computers. Computers process images in "digital" form, that is, as collections of digits (numbers). A typical black-and-white digital image consists of thousands or millions of numbers laid out in a rectangular array like the squares on a checkered tablecloth. (The numbers are not stored this way physically in the computer, but they are organized as if they were.) To turn this array of numbers into a visible image, as when making a printout or displaying the image on a screen, a tiny, visible dot is created from each number. Each dot is called a picture element or "pixel." A color image of the same size consists of three times as many numbers as a black-and-white image because there are three numbers per pixel, one number for the brightness of each color channel. The three colors used may be the three primary colors (red, yellow, blue), the three secondary colors (cyan, magenta, yellow), or the colors of the popular RGB scheme (red, green, blue). By adding different amounts from each color channel, using the three numbers for each pixel as a recipe, a pixel of any color can be made.

A rectangular array of numbers is also called a "matrix." An entire field of mathematics—"matrix algebra"—is devoted to working with matrices. Matrix algebra may be used to change the appearance of a digital image, extract information from it, compare it to another image, merge it with another image, and to affect it in many other ways. The techniques of Fourier transforms, probability and statistics, correlation, wavelets, artificial intelligence, and many other fields of mathematics are applied to digital images in art, engineering, science, entertainment, industry, police work, sports, and warfare, with new methods being devised every year.

In general, we are interested in either creating, altering, or analyzing images.

A Brief History of Discovery and Development

The relationship between images and mathematics began with the invention of classical geometry by Greek thinkers such as Euclid (c. 300 B.C.) and by mathematicians of other ancient civilizations. Classical geometry describes the properties of regular shapes that can be drawn using curved and straight lines, namely, geometric figures such as circles, squares, and triangles and solids such as spheres, cubes, and tetrahedra. The extension of mathematics to many types of images, not just geometric figures, began with the invention of perspective in the early 1400s. Perspective is the art of drawing or painting things so as to create an illusion of depth. In a perspective drawing, things that are farther from the artist are smaller and closer together according to strict geometric rules. Perspective became possible when people realized that they could apply geometry to the space in a picture, rather than just to shapes such as circles and triangles. Today, the mathematics of perspective—specifically, the group of geometric methods known as trigonometry—are basic to the creation of three-dimensional animations such as those in popular movies like *Jurassic Park* (1993), *Shrek* (2001), and *Star Wars Episode II: Attack of the Clones* (2003).

Real-life Applications

CREATING IMAGES

Because a digital image is really a rectangular array matrix full of numbers, we can create one by inserting numbers into a matrix. This is done, most often in the movie industry, by cooking up numbers using mathematical tools such as Euclidean geometry, optics, and fractals. A digital image can also be created by scanning or digitally photographing an existing object or scene.

ALTERING IMAGES

The most common way of altering a digital image is to take the numbers that make it up and apply some mathematical rule to them to create a new image. Methods of this kind including enhancement (making an image look better), filtering (removing or enhancing certain features of the image, like sharp edges), restoration (undoing damage like dust, rips, stains, and lost pixels), geometric transformation (changing the shape or orientation of an image), and compression (recording an image using fewer numbers). Most home computers today contain software for doing all these things to digital images.

Sports Video Analysis

Video analysis is the use of mathematical techniques from probability, graph theory, geometry, and other areas to analyze sports and other kinds of videos. Sports video analysis is a particularly large market, with millions of avid watchers keen for instant replays and new and better ways of seeing the game.

Traditionally, the only way to find specific moments in a video (or any other kind) of video was to fast-forward through the whole thing, which is time-consuming and annoying. Today, however, mathematics applied to game footage by computers can automatically locate specific plays, shots, or other moments in a game. It can track the ball and specific players, automatically extract highlights and statistics, and provide computer-assisted refereeing. Soon, three-dimensional computer models of the game space constructed from multiple cameras will allow the viewer to choose their own viewpoint from which to view the game as if from the front row, floating above the field, following a certain player, following the ball, or wherever. Some software based on these techniques, such as the Hawk-Eye program used to track the ball in broadcast cricket matches, is already in commercial use.

Video analysis in sports is also used by coaches and athletes to improve performance. Mathematical video analysis can show exactly how a shot-putter has thrown a shot, or how well the members of a crew team are pulling. By combining global positioning system (GPS) information about team players' exact movements with computerized video analysis and radio-transmitted information about breathing and heart rates, coaches (well-funded, high-tech, and "math savvy" coaches, that is) can now get an exact picture of overall team effort.

ANALYZING IMAGES

Analyzing an image usually means identifying the objects in it. Is that blob a face, a potato, or a bomb in the luggage? If it's a face, whose face is it? Is that dark patch in the satellite photograph a city, a lake, or a plowed field? Such questions are answered using a wide array of mathematical techniques that reduce images to representation of pixels by numbers that are then subject to mathematical analysis and operations.

OPTICS

Mathematics and imaging formed another fruitful connection with the growth of modern mathematical optics starting in the 1200s. Mathematical optics is the study of images are formed by light reflecting from curved mirrors or passing through one or more lenses and falling on any flat or light-sensitive surface such the retina of the eye, a piece of photographic film, or a light-sensitive circuit such as is used in today's digital cameras. Mathematical optics makes possible the design of contacts, eyeglasses, telescopes, microscopes, and cameras of all kinds. Advanced mathematics are needed to predict the course of light rays passing through many pieces of glass in high-quality camera lenses, and to design lens shapes and coatings that will deliver a nearly perfect image.

MEDICAL IMAGING

For the better part of a century, starting in the 1890s, the only way to see anything inside of a human body without cutting it open was to shine x rays through it. Shadows of bones and other objects in the body would cast by the x rays on a piece of photographic film placed on the other side of the body. This had the disadvantages that it could not take pictures of soft tissues deep in the body (because they cast such faint shadows), and that the shadows of objects in the path of the x-ray beam were confusingly overlaid on the x-ray film. Further, excessive x-ray doses can cause cancer. However, the spread of inexpensive computer power since the 1960s has led to an explosion of medical imaging methods.

Due in part to faster computers, it is now possible to produce images from x-rays and other forms of energy, including radio waves and electrical currents, that pass through the body from many different directions. By applying advanced mathematics to these signals, it is possible to piece together extremely clear images of the inside of the body—including the soft tissues. Magnetic resonance imaging (MRI), which places the body in a strong magnetic field and bombards it with radio waves, is now widely available. A technique called "functional MRI" allows neurologists to watch chemical changes in the living brain in real time, showing what parts of the brain are involved in thinking what kinds of thoughts. This has greatly advanced our knowledge of such brain diseases as Alzheimer disease, epilepsy, dyslexia, and schizophrenia.

COMPRESSION

Imagine a square digital image 1,000 pixels wide by 1,000 pixels tall—all one solid color, blue. That's 1,000 × 1,000 or 1 million blue pixels. If each pixel requires 3 bytes (one byte equals eight bits, that is, eight 1s and 0s), this extremely dull picture will take up 3 million bytes (megabytes, MB) of computer memory. But we don't need to waste 3 MB of memory on a blue square, or wait while they transmit over the Web. We could just say "blue square, 1,000 pixels wide" and have done with it: everything there is to know about that picture is summed up by that phrase. This is an example of "image compression." Image compression takes advantage of the redundancy in images—the fact that nearby pixels are often similar—to reduce the amount of data storage and transmission time taken up by images. Many mathematical techniques of image compression have been developed, for use in everything from space probes to home computers, but the most of the images that are received and sent over the World Wide Web are compressed by a standard method called JPEG, short for Joint Photographic Experts Group, first advanced in 1994.

JPEG is a "block encoding" method. This means that it divides the image up into blocks 8 by 8 pixels in size, then records as much of the image redundancy in that block as it can in a series of numbers called "coefficients." The coefficients that don't record as much redundancy are thrown away. This allows a smaller group of numbers (the coefficients that are left) to record most of the information that was in the original image. An image can then be reconstructed from the remaining coefficients. It is not quite as sharp as the original, but the difference may be too slight for the eye to notice.

RECOGNIZING FACES: A CONTROVERSIAL APPLICATION

Human beings are expert at recognizing faces. We effortlessly correct for different conditions of light and shadow, angles of view, glasses, and even aging. It is difficult, however, to teach a computer how to do this. Some progress has been made and a number of face-recognition systems are on the market.

The mathematics of face recognition are complex because faces do not always look the same. We can grow beards or long hair, don sunglasses, gain or lose weight, put on hats or heavy makeup, be photographed from different angles and in different lights, and age. To recognize a face it is therefore not enough to just look for matching patterns of image dots. A mathematical model of whatever it is that people recognize in a face—what it is about a face that doesn't change—must be constructed, if possible. Face-recognition software has a low success rate in real-life settings such as streets and airports, often wrongly matching people in the crowd with faces in the records or failing to identify people in the records who are in the crowd.

Face on Mars

In 1976, two spidery robots, *Viking 1* and *Viking 2*, became the first spacecraft to successfully touch down on the rocky soil of Mars. Each lander had a partner, an "orbiter" circling the planet and taking pictures. Images and other data from all four machines were radioed back to Earth.

One picture drew public attention from the first. It had been taken from space by a Viking orbiter, and it looked exactly like a giant, blurry face built into the surface of Mars (See Figure 1.)

Notice the dots sprinkled over the image. These are not black spots on Mars, but places where the radio signal transferring the image from the Viking orbiter as a series of numbers was destroyed by noise. However, one dot lands on the "nose" of the Face, right where a nostril would be; one lands on the chin, looking like the shadow of a lower lip; and several land in a curve more or less where a hairline would be. These accidents made the image look even more like a face.

Some people erroneously decided that an ancient civilization had been discovered on Mars. Scientists insisted that the "face" was a mountain, but a better picture was needed to resolve any doubt. In 2001 an orbiter with an better camera than Viking's did arrive at Mars, and it took the higher resolution picture of the "face" shown in Figure 2.

In this picture, the "Face" is clearly a natural feature with no particular resemblance to a human face. Thanks to mathematical processing of multiple images, we can now even view it in 3-D.

In later releases of Viking orbiter images in the 1970s the missing-data dots were "interpolated," that is, filled in with brightness values guessed by averaging surrounding pixels. Without its dots, and seen in more realistic detail, the "Face" does not look so face-like after all.

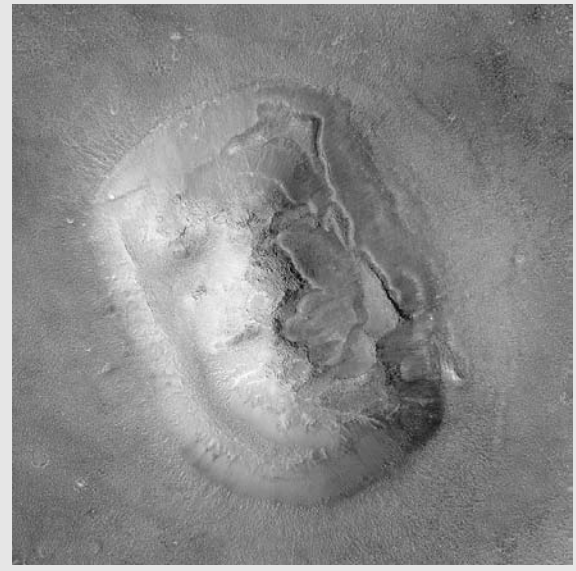


Figure 1 (top). NASA/JPL/MSSS.
Figure 2 (bottom). 1989 ROGER RESSMEYER/NASA/CORBIS.

Using face-recognition systems to scan public spaces is politically controversial. At the Super Bowl game in Tampa, Florida, in 2001, for example, officials set up cameras to scan the fans as they went through the turnstiles. The videos were analyzed using face-recognition software. A couple of ticket scalpers were caught, but no serious criminals. Face-recognition technology has not been used again at a mass sporting event, but is in use at several major airports, including those in Boston, San Francisco, and Providence, Rhode Island.

Critics argue that officials might eventually be able to track any person's movements automatically, using the thousands of surveillance cameras that are being installed to watch public spaces across the country. Such a technology could be used not only to catch terrorists (if we knew what they looked like) but, conceivably, to track people for other reasons.

Face-recognition systems may prove more useful and less controversial in less public settings. Your own computer—which always sees you from about the same angle, and in similar lighting—may soon be able to check your identity before allowing you to spend money or access secure files. Some gambling casinos already use face-recognition software to verify the identities of people withdrawing winnings from automatic banking machines.

FORENSIC DIGITAL IMAGING: SHOEPRINTS AND FINGERPRINTS

Forensic digital imaging is the analysis of digital images for crime-solving. It includes using computers to decide whether documents are real or fake, or even whether the print of a shoe at a crime scene belongs to a particular shoe. Shoeprints, which have been used in crime detection even longer than fingerprints, are routinely photographed at crime scenes. These images are stored in large databases because police would like to know whether a given shoe has appeared at more than one crime scene. Matching shoe prints has traditionally been done by eye, but this is tedious, time-consuming, and prone to mistakes. Systems are now being developed that apply mathematical techniques such as fractal decomposition to the matching of fresh shoeprints with database images—faster and more accurately than a human expert. Fingerprints, too, are now being translated into digital images and subjected to mathematical analysis. Evidence that will stand up in court can sometimes now be extracted from fingerprints that human experts pronounced useless years ago.

DANCE

Dance and other motions of the human body can be described mathematically. This knowledge can then be used to produce computer animations or to record the choreography of a certain dance. In Japan, for example, the number of people who know how to dance in traditional style has been slowly decreasing. Some movies and videos, however, have been taken of the older dances. Researchers have applied mathematical techniques to these videos—some of which have deteriorated from age and are not easy to view—in order to extract the most complete possible description of the various dances. It would be better if the dances could be passed down from person to person, as they have in the past, but at least in this way they will not be completely forgotten. Japanese researchers, who are particularly interested in developing human-shaped robots, also hope to use mathematical descriptions of human motion to teach robots how to sit, stand, walk—and dance.

MEAT AND POTATOES

The current United States beef-grading system assigns a grade or rank to different pieces of beef based on how much fat they contain (marbling). Until recently, an animal had to be butchered and its meat looked at by a human inspector in order to decide how marbled it was. However, computer analysis of ultrasound images has made it possible to grade meat on the hoof—while the animal is still alive. Ultrasound is any sound too high for the ear to hear. It can be beamed painlessly into the body of a cow (or person). When this is done, some of the sound is reflected back by the muscles and other tissues in the body. These echoes can be recorded and turned into images. In medicine, ultrasound images can reveal the health of a human fetus; in agriculture, mathematical techniques like gray-scale statistical analysis, gray-scale spatial texture analysis, and frequency spectrum texture analysis can be applied to them in order to decide the degree of marbling.

Different mathematics are applied to the sorting of another food item that often appears at mealtime with meat: potatoes. Potatoes that are the right size and shape for baking can be sold for higher price, and so it is desirable to sort these out. This can either be done hand or by passing them down a conveyer belt under a camera connected to a computer. The computer is programmed to decide which blobs in the image are potatoes, how big each potato is, and whether the potatoes that are big enough for baking are also the right shape. All these steps involve imaging mathematics.

STEGANOGRAPHY AND DIGITAL WATERMARKS

For thousands of years, people have been interested in the art of secret messages (also called "cryptography," from

Key Terms

Matrix: A rectangular array of variables or numbers, often shown with square brackets enclosing the array. Here "rectangular" means composed of columns of equal length, not two-dimensional. A matrix equation can represent a system of linear equations.

Pixel: Short for "picture unit," a pixel is the smallest unit of a computer graphic or image. It is also represented as a binary number.

the Greek words for "secret writing"), and computers have now made cryptography a part of everyday life; for example, every time someone uses a credit card to buy something over the Internet, their computer uses a secret code to keep their card number from being stolen. The writing and reading of cryptographic or secret messages by computer is a mathematical process.

But for every code there is a would-be code-breaker, somebody who wants to read the secret message. (If there wasn't, why would the message be secret?) And a message that looks like it is in a secret code—a random-looking string of letters or numbers—is bound to attract the attention of a code-breaker. Your message would be even more secure if you could keep its very existence a secret. This is done by steganography (from the Greek for "covered writing"), the hiding of secret messages inside other messages, "carrier" messages, that do not appear secret at all. Secret messages can be hidden physically (a tiny negative under a postal stamp, or disguised as a punctuation mark in a printed letter) or mathematically, as part of a message coded in letters, numbers, or DNA. Digital images are particularly popular carriers. We send many images to each other, and an image always has an obvious message of its own; by drawing attention to itself, an image diverts suspicion from itself. But a digital image may be much more than it appears. The matrix of numbers that makes it up can be altered slightly by mathematical algorithms to convey a message while changing the visible appearance of the image very little, or not at all. And since images contain so much more binary information than texts such as letters, it is easier to hide longer secret messages in them.

You do not have to be a spy to want to hide a message in an image. People who copyright digital photographs want to prevent other people from copying them and using them for free, without permission; one way to do so is to code a hidden owner's mark, a "digital watermark," into the image. Software exists that scans the Web looking for images containing these digital watermarks and checking to see whether they are being used without permission.

ART

Digital imaging and the application of mathematics to digital images have proved important to the caretaking of a kind of images that are emphatically not digital, not a mass of numbers floating in cyberspace, not reproducible by mere copying of 1s and 0s: paintings of the sort that hang in museums and collections. Unlike digital images, these are physical objects with a definite and unique history. They cannot be truly copied and may often be worth many millions of dollars apiece. The role of digital imaging is not to replace such paintings, but to aid in their preservation.

The first step is to take a super-high-grade digital photograph of the painting. This is done using special cameras that record color in seven color bands (rather than the usually three) and take extremely detailed scans. For example, a fine-art scanner may create a digital image $20{,}000 \times 20{,}000$ pixels (color dots) large, which is 400 million pixels total. But each pixel has seven color bands, so there are actually seven times this many numbers in the image record, about 2.8 billion numbers per painting. This is about 100 times larger than the image created by a high-quality handheld digital camera.

Once this high-grade image exists, it has many uses. Even in the cleanest museum, paintings slowly dim, age, and get dirty, and so must eventually be cleaned up or "restored." A digital image shows exactly what a painting looks like on the day it was scanned; by re-scanning the painting years later and comparing the old and new images using mathematical algorithms, any subtle changes can be caught. By applying mathematical transformations to the image of a painting whose colors have faded, experts can, in effect, look back in time to what the painting used to look like (probably), or predict what it will look like after cleaning. Also, famous paintings are often transported around the world to show in different museum. By re-imaging a painting before and after transport and comparing the images, any damage during transport can be detected.

Where to Learn More

Web sites

"Computer Technology in Soccer." Soccer Performance.org. <http://www.soccerperformance.org/specialtopics/companalsystems.htm> (October 16, 2004).

Johnson, N.F. "Steganography and Digital Watermarking." 2002. <http://www.jjtc.com/Steganography/> (October 16, 2004).

"Privacy and Technology: Q&A on Face-Recognition." American Civil Liberties Union. Sep. 2, 2003. <http://www.aclu.org/Privacy/Privacy.cfm?ID=13434&c=130> (October 16, 2004).

Kimmel, R., and G. Sapiro. "The Mathematics of Face Recognition." Society for Industrial and Applied Mathematics (SIAM). SIAM News, Volume 36, Number 3, April 2003. <http://www.siam.org/siamnews/04-03/face.htm> (October 16, 2004).

Wang, J.R., and N. Parameswaran. "Survey of Sports Video Analysis: Research Issues and Applications." Australian Computer Society. Conferences in Research and Practice in Information Technology, Vol. 36. M. Piccardi, T. Hintz, X. He, M.L. Huang, D.D. Feng, J. Jin, Eds. 2004. <http://crpit.com/confpapers/CRPITV36Wang.pdf> (October 16, 2004).

Information Theory

Overview

It is often said that we live in the Information Age. Computer enthusiasts sometimes speak as if we were now being fed and housed by the "information economy," or as if we were all racing down the "information highway" toward a perfect society. But what, exactly, is "information"? We all know that disks and chips store it, and that computers process it, and that is supposed to be a good thing to have lots of—but what is it?

The answer is given by information theory, a branch of mathematics founded in 1948 by American telephone engineer Claude Shannon (1916-2001). Shannon discovered how to measure the amount of information in any given message. He also showed how to measure the ability of any information-carrying channel to transmit information in the presence of noise (which disrupts and changes messages). Information theory soon expanded to include error-correction coding, the science of transmitting messages with the fewest possible mistakes.

Shannon's ideas about information have proved useful for many things besides telephones. Information theory enables designers to make many kinds of message-handling devices more efficient, including compact disc (CD) players, deep-space probes, computer memories, and other gadgets. Information theory has also proved useful in biology, where the DNA molecules that help to shape us from birth to death turn out to be written in code, and in economics, where information processing is key to making money in a complicated, competitive world. Error-correction coding also enables billions of files to be transferred over the Internet every day with few errors.

Fundamental Mathematical Concepts and Terms

The central idea of information theory is information itself. In everyday speech, "information" is used to mean "useful knowledge"; if you have information about something, you know something useful or significant about that thing. In mathematics, however, the word has a much narrower meaning.

Shannon began with the simple idea that whatever information is, messages carry it. From this he derived a precise mathematical expression for the information in any given message. Every system that transmits a message has, Shannon said, three parts: a sender, a channel, and a receiver. If the sender is a talker on one end of a phone line, the phone line is the channel and the listener at the far end is the receiver.

We imagine that the sender chooses messages from a collection of possible messages and sends them one by one through the channel. Say that N stands for the number of messages that the sender has to choose from each time. If the message is a word from the English language, N is about 600,000. Often the message is a string of ones and zeroes. Ones and zeroes are often used to represent messages because they are easy to handle. Each one or zero is called a "binary digit" (or "bit," for short). If a binary message is M bits long, then the number of possible messages, N, equals 2^M. This is because there can be only 2^M different strings of ones and zeroes M digits long. For example, if the message could be any string 3 bits long ($N = 3$), then $M = 8$, because there are $2^3 = 8$ different 3-bit strings, as shown in Table 1.

000	101
100	110
010	111
001	011

Table 1.

The sender chooses a message at random. Here, random means that all N messages are equally likely, just as, when you flip a fair coin, heads and tails are equally likely. If all N messages are equally likely, the chance or probability of each message being sent is $1/N$. For example, if we flip a coin to choose whether to send 1 or 0 (1 for heads, 0 for tails), then $N = 2$ (there are two possible messages) and the probability of each message is 1/2 (because $1/N = 1/2$).

From the sender's point of view, the situation is simple: choose a message and send it. From the receiver's point of view, things are less simple. The receiver knows that a message is coming, but they do not know which one. They are therefore said to have uncertainty about what message will be sent. Exactly how much "uncertainty" they have depends on N. That is, the more possible messages there are (the larger N is), the harder it is for the receiver to guess what message will be sent.

The receiver's uncertainty is important because it tells us exactly much they learn by receiving a message. If there is only one possible message—say, if the sender can only send the digit "0", over and over— then the receiver can always "guess" it ahead of time, so they learn nothing by receiving it. If there are two possible messages ($N = 2$), then the receiver has only a 50–50 chance of guessing which will be sent, and definitely learns something when a message is received. If there are more than two possible messages ($N > 2$), then the receiver's chance of guessing which message will be sent is less than 50–50.

The harder it is to guess a message before getting it, the more one learns by getting it. Therefore, the receiver's uncertainty tells us how much they learn—how much information they gain—from each message. Messages chosen at random from large message-sets are harder to guess ahead of time, so the receiver learns more by receiving them; they convey more "information."

Now assume that a message has been chosen from the list of N possibilities, sent, and correctly received. The receiver's uncertainty about this particular message has now been reduced to 0. This reduction in uncertainty corresponds, as we have seen, to a gain in information. This, then, is information theory's definition of information: Information is what reduces uncertainty. We will label information H, as is customary.

The information that the receiver derives from a single message, H, depends on the number of possible messages, N. Bigger N means more uncertainty: more uncertainty means more information gained when the message arrives. To signify the dependence of information on N, we write H as a "function" of N, like so: $H(N)$. (This is pronounced "H of N".) A function is a rule that relates one set of numbers to another set. For example, if we write $f(x)$, we mean that for every number x there is another number, f, related to it by some rule; if the rule is, for example, that f is always twice x, we write $f(x) = 2x$. Likewise when we write $H(N)$, we say that for every N there is another number, H, related to it by some rule. Below, we'll look at exactly what this rule is.

H, which stands for the amount of information in a single message, has units of "bits." Similarly, numbers that record distances have units of feet (or meters, or miles) and numbers that record time intervals have units of seconds (or hours, or days). The bit is defined as follows: If a message consisting of a single binary digit is received, and that message was equally likely to be a 1 or a 0, then 1 bit of information has been received.

To find out what the function or rule is that relates the numbers H and N, we first introduce an imaginary wrinkle. Let us say that the sender is picking messages from two groups of possibilities, like two buckets of marbles. One group of possible messages has N_1 choices (Bucket Number 1, with N_1 marbles in it) and the other has N_2 choices (Bucket Number 2, with N_2 marbles in it). (The small "1" and small "2" attached to N_1 and N_2 are just labels that help us tell the two numbers apart.) Now imagine that the sender picks a message from the first group and sends it, then picks a message from the second

group and sends it, like grabbing one marble from Bucket Number 1 and a second marble from Bucket Number 2.

This sender is really sending messages (or picking marbles) in *pairs*. How many such pairs could there be? If we call the number of possible pairs N, then $N = N_1N_2$. This is easy to see with simple groups of messages. If the first message is a 0 or 1 (a single binary digit), then $N_1 = 2$, and if the second set is a *pair* of binary digits, the four possible messages are 00, 10, 01, and 11, so $N_2 = 4$. Choosing one message from each set allows eight (that is, $N_1 \times N_2$) possible pairs, as shown in Figure 1.

Message 1	Message 2	Eight possible message pairs
0	00	0 00
0	10	0 10
0	01	0 01
0	11	0 11
1	00	1 00
1	10	1 10
1	01	1 01
1	11	1 11

Figure 1.

It is easy to prove to yourself that these really are the only possible message pairs—just try to write one down that isn't already on the list.

How much information does one of these message-pairs contain? To give a specific number we would have to know the correct rule for relating H and N, that is, the function $H(N)$, which is what we're still looking for. But we can say one thing right off the bat: $H(N)$ should agree with common sense that the information given by the two messages together is the sum of the information given by the two messages separately. It turns out that this common-sense idea is the key to finding $H(N)$. Saying that the information in the two messages adds can be written as follows: $H(N) = H(N_1) + H(N_2)$.

But we also know, as shown above, that $N = N_1 \times N_2$. We can therefore rewrite the previous equation a little differently: $H(N_1N_2) = H(N_1) + H(N_2)$. It may not seem like we've proven much by writing this equation, but it is actually the key to our whole problem. Because it has the form it has, there is only one possible way to compute the information content of a message, that is, one possible rule or function. Mathematicians have shown, using techniques too advanced to go over here, that there is only one function that satisfies $H(N_1N_2) = H(N_1) + H(N_2)$, namely $H(N) = \log_2 N$.

The expression "$\log_2 N$" means "the base 2 logarithm of N," namely, that power of 2 which gives N. For example, $\log_2 8 = 3$ because $2^3 = 8$, and $\log_2 = 4$ because $2^4 = 16$. (See the entry in this book on Logarithms.) The graph of $H(N) = \log_2 N$ is shown in Figure 2.

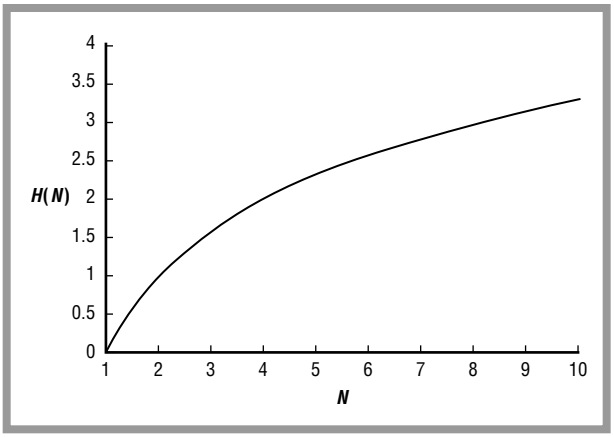


Figure 2. The information content of a single message selected from N equally likely messages: $H(N) = \log_2 N$. Units of $H(N)$ are bits.

As we saw earlier, the number of different messages that can be sent using binary digits (ones and zeroes) is $N = 2^M$. So, for example, if we send messages consisting of 7 binary digits apiece, the number of different messages is $N = 2^7 = 128$. Using 2^M for N in $H(N) = \log_2 N$, we get a new expression for $H(N)$: $H(M) = \log_2 2^M$.

But $\log_2 2^M$ is just M, by the definition of the base-2 logarithm given above, so $H(M) = \log_2 2^M$ simplifies to $H(M) = M$. This is just a straight line, the simplest of all functions, as shown in Figure 3. The equation $H(M) = M$ not only looks simple, it has a simple meaning: a message written using M equally likely binary digits conveys M bits of information. This is why we use "bit," an abbreviation of "binary digit," as the unit of information.

It is important to remember that while the "bit" is the unit of all information, not all information is in the form of "binary digits" (e.g., ones and zeroes). For example, the letters in this sentence are not binary digits, but they contain information.

UNEQUALLY LIKELY MESSAGES

A bit is the amount of information conveyed by the answer to the simplest possible question, that is, a question with two equally likely answers. When a lawyer in a courtroom drama shrieks "Answer yes or no!" at a witness, they are asking for one bit of information. Paul Revere's famous scheme for anticipating a British raid

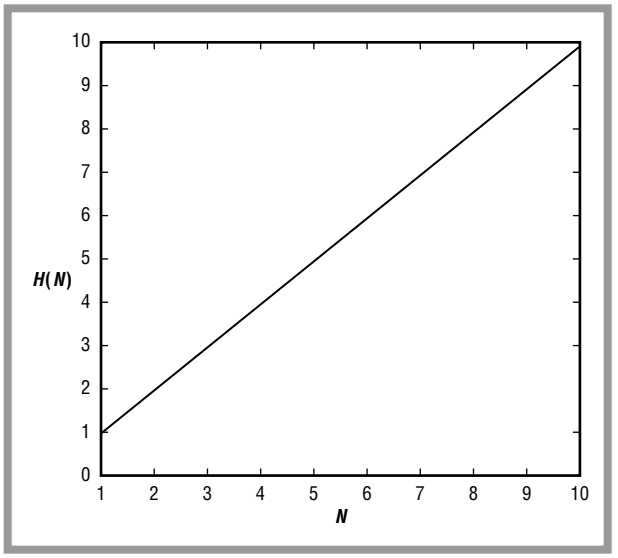


Figure 3. Bits of information, *H*, in a message, shown as a function of the number of binary digits in that message, *N*.

from Boston, as described by Henry Longfellow (1807–1882) in the famous poem beginning "Listen my children, and you shall hear / Of the midnight ride of Paul Revere," sought to convey one bit of information:

> [Revere] said to his friend, "If the British
> march
> By land or sea from the town tonight,
> Hang a lantern aloft in the belfry arch
> Of the North Church tower as a signal light,—
> One, if by land, and two, if by sea;
> And I on the opposite shore will be,
> Ready to ride and spread the alarm . . ."

Strictly speaking, this was a one-bit message only if the British were equally likely to come by land or by sea.

But what if they were not? What if the British were, say, five times as likely to come by land as by sea? So far we've talked about messages selected from equally likely choices, but what if the choices aren't equally likely?

In that case, our rule for the information content of a message must become more complicated. It also becomes more useful, because it is usually the case that some messages are more likely than others. In transmitting written English, for example, not all letters of the alphabet are equally likely; we send the letter "e" about 1.36 times more often than the next most common letter, "i."

Let's say that the sender has three messages to choose from, only now each message has a different chance or probability of being sent. The probability of an event is written as a number between 0 and 1: smaller probability numbers mean less-likely events, larger numbers mean more-likely events. Say that the probability of the first message on the sender's list is p_1, that of the second message is p_2, and that of the third message is f^3. The amount of information per message is, in this case, given by the following equation: $H(N) = -p_1\log_2 p_1 - p_2\log_2 p_2 - p_3\log_2 p_3$ bits.

If there were more than 3 possible messages, there would be more terms to subtract, such as $p_4\log_2 p_4$, $p_5\log_2 p_5$, and so on up to as many terms as there were possible messages.

These equations are the heart and soul of information theory. Using it, we can calculate exactly how much information, *H*, any message is worth, if we know the probabilities of all the possible messages. This is best explained by working out a simple example.

Paul Revere had two possible messages to deliver, "land" or "sea," so in his case $N = 2$. We will call p_1 the probability that the message would be "land," and p_2 the probability that it would be "sea". In this case then, $H(N) = -p_1\log_2 p_1 - p_2\log_2 p_2$ bits. If both messages are equally likely, then p_1 and p_2 both equal 1/2 and so we have

$$H(2) = -\left(\frac{1}{2}\log_2\frac{1}{2}\right) - \left(\frac{1}{2}\log_2\frac{1}{2}\right)\text{bits}$$

which works out to $H(2) = 1$ bit. This, we already knew: Where there are two equally likely messages, sending either one communicates 1 bit of information.

But if the probabilities of the two messages are not equal, less than 1 bit is communicated. For example, if $p_1 = .7$ and $p_2 = .3$ then $H(2) = -(.7\log_2 .7) - (.3\log_2 .3) = .88129$ bits.

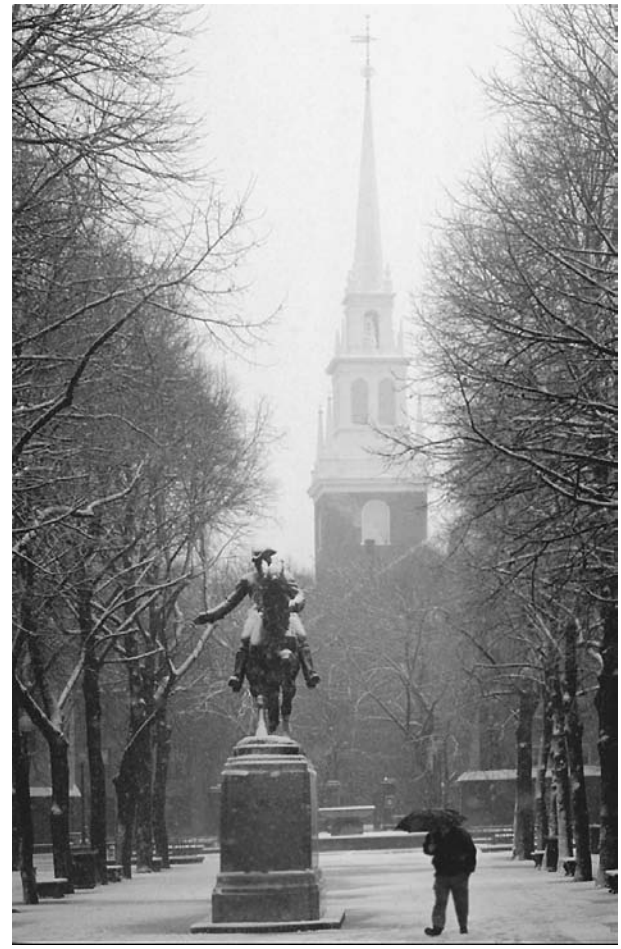
This agrees with common sense, which tells us that if the American revolutionaries had known beforehand that the British were more than twice as likely to come by land than by sea (probability .7 for land, only .3 for sea), they would have had a pretty good shot at guessing what was going to happen even without getting the message from the church tower (and so that message wouldn't have told them as much as in the equal-probability case). If the revolutionaries had known that the British were sure to come by land, then p_1 would have equaled 1 (the probability of a certain event), p_2 would have equaled 0 (the probability of an impossible event), and the message would have communicated no information, zero bits: $H(2) = -(1 \times \log_2 1) - (0 \times \log_2 0) = 0$ bits.

And that makes sense too. A message that conveys 0 bits is one that you don't need to receive at all.

INFORMATION AND MEANING

The assignment of Revere's colleague in the church tower was to send a single binary digit: "one, if by land, and two, if by sea." If the person in the tower had written "Land" or "Sea" on paper, instead of putting up lights, the message would have contained more bits of information—about 18.8 bits for "Land" and 14.1 bits for "Sea," taking each letter as worth $\log_2 26 = 4.7$ bits (because there are 26 letters in the alphabet)—yet the message would have *meant*—the same thing. This seems like a contradiction: More information does not necessarily provide greater knowledge. Why not?

The answer is that the everyday sense of the word "information" is different from the mathematical sense. The everyday sense is based on meaning or importance. If a message is meaningful, that is, tells us something important, we tend to think of it as having more information in it. Mathematically, however, this isn't true. How much information a message contains has nothing to do with how meaningful that message is. The answers to 1-bit, "Yes-No" questions like "Shall we surrender?" or "Will you marry me?", which are very important, contain only 1 bit of information. On the other hand, a many-bit message, say a hundred 1s and 0s picked by flipping a coin, may have no meaning at all. Meaning and information are not the same thing.



The Paul Revere statue in Paul Revere Plaza in the North End neighborhood of Boston. The spire of the famous Old North Church is seen in the background. According to information theory, Paul Revere's famous scheme for anticipating a British raid from Boston, as described by Henry Longfellow in the famous poem beginning "Listen my children, and you shall hear / Of the midnight ride of Paul Revere," sought to convey approximately 1 bit of information. AP/WIDE WORLD PHOTOS. REPRODUCED BY PERMISSION.

A Brief History of Discovery and Development

Information theory dates from the publication of Claude Shannon's 1948 paper, "A Mathematical Theory of Communication." A few scientists had suggested using a logarithmic measure of information before this, but Shannon—who was famous for riding a unicycle up and down the hallways of Bell Laboratories—was the first to hit on the necessary mathematical expressions. He defined "information," distinguished it from meaning, and proved several important theorems about transmitting it in the presence of noise (random signals that cause erroneous messages to be received).

Real-life Applications

A great deal of work has been done on information theory since Shannon's 1948 paper, applying and extending his ideas in thousands of ways. Few of us go through a single day without availing ourselves of some application of information theory. Cell phones, MP3 players, palm pilots, global positioning system units, and laptops all rely information theory to operate efficiently.). Information theory is also applied to electronic communications, computing, biology, linguistics, business, cryptography, psychology, and physics. It an essential branch of the mathematical theory of probability.

COMMUNICATIONS

Paul Revere was not only a revolutionary conspirator, but part of a communications channel. Once he had seen whether one light or two was burning in the church tower, it was his job to deliver that one precious bit of information to its final destination, the revolutionary militia at Concord, Massachusetts.

However, he was captured by a British patrol before getting there. He was thus part of what engineers call a "noisy channel." All real-world channels are noisy, that is, there is some chance, large or small, that every message will suffer damage or loss before it gets to its intended receiver. Before Shannon, engineers mostly thought that the only way to guarantee transmission through a noisy channel was to send messages more slowly. However, Shannon proved that this was wrong. Every channel has a certain capacity, that is, a rate at which it can send information with as few errors as you please *if* you are allowed to send a certain number of extra, redundant bits—information that repeats other parts of your message—along with your actual message.

Shannon showed how to calculate channel capacity exactly. With this tool in hand, engineers have known for half a century how to make every message channel as good as it needs to be, squeezing the most work possible out of every communications device in our increasingly gadget-dependent world: optical disc drives, cell phones, optical fibers carrying hundreds of thousands of telephone calls through underground pipes, radio links with deep-space probes, file transfers over the Internet, and so on.

Around A.D. 1200, the Chinese were able to invent primitive rockets without knowing calculus or Newton's Laws of Motion, but without mathematics they could never have built truly huge rockets such as the Long March 2F booster that lifted the first Chinese astronaut into space in May 2004. Likewise, communications devices and digital computers were invented before information theory, but without information theory engineers could not build such machines as well (and as cheaply) as we do today. In rocketry, communications, powered flight, and many other fields, the early steps depend mostly on creative spunk but later improvements depend on mathematics.

Today applications of information theory are literally everywhere. Every cubic inch of your body is at this moment interpenetrated by scores or hundreds of radio signals designed using information theory.

Physics and Information

From the very beginning there has been a connection between physics and information theory. Shannon's rule for calculating information was nearly identical to the expression in statistical physics for the entropy of a system (a measure of its disorder or randomness), as was pointed out to Shannon before he published his famous 1948 paper. One physicist even advised Shannon to call his new measure "entropy," not "information," because "most people don't know what 'entropy' really is. If you use 'entropy' in an argument you will win every time!"

Nor is the connection between physics and information merely a matter of look-alike equations. In 1951, the physicist L. Brillouin proved the amazing claim that there is an absolute lower limit on how much energy it takes to observe a single bit of information. Namely, to observe one bit takes at least $kBT\log_e 2$ ergs of energy, where T is temperature in degrees Kelvin, k and B are constants (fixed numbers) from physics, and $\log_e 2$ equals approximately .693. The precise value of this very small number is not important: what is important is what it tells us. One of the things it tells us that it is impossible to have or process an infinite amount of information. That would, by Brillouin's theorem, take an infinite amount of energy; but there is only a limited amount of energy in the whole Universe.

INFORMATION THEORY IN BIOLOGY AND GENETICS

Most of the cells in your body contain molecules of DNA (deoxyribonucleic acid). Each DNA molecule is shaped like a long, narrow ribbon or zipper that has been twisted lengthwise like a licorice stick. Each side of the zipper has a row of teeth, each tooth being a cluster of atoms. There are four kinds of zipper teeth in DNA, the chemicals guanine, thymine, adenine, and cytosine (always called G, T, A, and C for short).

These teeth of the DNA zipper are lined up in groups of three: AGC, GGT, TCA, and so on for many thousands of groups. Each group of three teeth is a code word bearing a definite message. Also, each type of zipper tooth is shaped so that it can link up with only one other kind of zipper tooth: A and T always zip together, and G and C always zip together. Therefore, both sides of the zipper bear the same series of messages, only coded with different chemicals: thus, AGCGGT zips together with TCGCCA. If you know what one side of the zipper looks like, you can say what the other side must look like.

DNA is usually zipped up so that the two opposite sets of teeth are locked together. Sometimes, however, DNA gets partly unzipped. This happens whenever the cell needs to read off some of the messages in the DNA, such as when the cell needs to make a copy of itself or to refresh its stores of some useful chemical. The unzipping is done by special molecules that move down the DNA, separating the two sides like the slide on an actual zipper. When a section of DNA has been unzipped, other molecules move along it and copy (or "transcribe") its three-letter code words. These code words order the cell to

string certain molecules ("amino acids") together like beads on a necklace. These strung-together amino acids are the very complex molecules called proteins, which do most of the microscopic, chemical work that keeps us alive. Proteins are produced from step-by-step instructions in DNA much as a cook bakes a cake from step-by-step instructions (a recipe) in a cookbook. The exact same three-letter DNA code is used in the cells of every living thing on Earth, from people to pine trees.

Biologists have found it helpful to view each three-letter DNA code word as a message. Since there are four choices of letter (A, C, G, and T) and three letters per word, there are $4^3 = 64$ possible words that DNA might send. According to information theory, each DNA code word could contain up to $\log_2 64 = 6$ bits of information. Actually, some words are used by DNA to mean the same thing as other words, so the DNA code only codes for 20 different amino acids, not 64. Each DNA code word therefore contains about $\log_2 20 = 4.32$ bits. There are about three billion pairs of molecular zipper teeth (base pairs) in a complete set of human DNA molecules. These three billion pairs could encode, at most, one billion three-letter words, each conveying 4.32 bits. Therefore, the most information that the human DNA could contain is about 4.32 billion bits. A standard 700 MB CD-ROM also contains about this much information.

Thus, an entire CD-ROM's worth of information is packed by Nature into a chemical speck too small to be seen without a powerful microscope—a set of human DNA molecules. Most of the cells in the human body contain these molecules.

Accordingly, the "recipe" for a human being requires about as much information storage space as it would take to record 80 minutes of dance hits.

Seeing the DNA-to-protein system in terms of information theory has helped biologists understand evolution, aging, growth, and viruses such as AIDS. Biologists have also applied information theory to molecules other than DNA and to the brain.

ERROR CORRECTION

Every message has some chance of not getting through or of getting through with damage, like a letter that is delivered with a corner torn off or with a letter "O" smeared into a letter "Q." Here is another problem that begs for a clever solution.

Once again, Paul Revere is ahead of us. Revere's task was to deliver his one-bit message to the town of Concord, Massachusetts. On the way there, he stopped at Lexington and shared the message with two other men, William Dawes and Samuel Prescott. All three set off for Concord; all three were captured by the British. Revere was released without his horse. Dawes and Prescott made a break for it, but Dawes fell off his horse. Only Prescott got through. If Revere had headed straight for Concord by himself, the message would never have been delivered. Sending the message three separate times, by three separate riders, an example of "triple redundancy," got this one-bit message through this very noisy channel.

Today we send messages using electrons and photons rather than horses, but triple redundancy (sending a message three times) is still an option. For instance, instead of 101, we can send 111000111. If this message is damaged by electronic noise (static), then the receiver will receive a different message, for example, 011000111. In this case, noise has changed the first bit from a 1 to a 0. By looking at the first three bits the receiver knows, first of all, that an error must have happened, because all three bits are not the same. Triple redundancy thus has the power of *error detection*. Second, the receiver can decide whether the first three bits were a 1 or a 0 in the original message since there are two 1's and only one 0; thus 011 decodes to 1, which is correct. Triple redundancy also, therefore, has the power of *error correction*. In particular, if no more than one bit out of every three is changed by noise, the entire message can still get through correctly. If we were to send triple-redundant messages forever, we could send an infinite number of bits despite an infinite number of errors, as long as the errors didn't happen too fast!

In practice this scheme isn't used because it would be wasteful. It forces us to send three times as many bits as there are in the original message, but there are only a few simple errors that it can find and fix. If two bits that are close to each other get flipped by noise, we can find the error but our fix may be wrong: for instance, if 111 gets changed to 001 or 010, we will know that an error has happened (because the three bits are not all the same, as they should be), but by majority vote we will decode the received word incorrectly to 0, rather than 1. Errors that are near each other, as in this example, are called "burst" errors.

There are several ways to handle burst errors. The simplest that is used in many real-world codes is termed "interleaving." Interleaving takes one chunk of a message and slips its bits between the bits of another chunk, like two halves of a deck of cards being shuffled together. For example, we may want to transmit the message 01. We first create the two triply redundant words 000 and 111, then interleave them to get 010101. If two bits right next to each other get changed anywhere in this six-bit string, our simple code can both detect and correct them. If the second and third bits, for instance, are both changed

during transmission, 010101 is turned into 001101. Deshuffling this (taking the first, third, and fifth bits first, then the second, fourth, and sixth bits) gives us 010011. By majority vote, the first three bits decode to 0 and the second three bits decode to 1—which is our original message, 01. This shows the power of combining repetition with interleaving.

If *three* bits in a row are flipped, this code will fail. Every code has its limits. Nevertheless, Shannon's channel capacity theorem guarantees that we can always drive the corrected error rate down to any specific level we want, for a channel with a given amount of noise, by adding redundant bits. In the interleaved code with repetition that we've been considering, for example, we could correct longer burst errors simply by repeating each bit more than three times (e.g., six, or 10, or a 1,000 times). More redundancy, and more interleaving, would offer more protection. This is a general property of all error-correcting codes: You can never get something for nothing—but you can get something for something, namely, reliability for redundancy.

Information theorists call a code "perfect" if its level of error correction is bought for the least possible redundancy. Error correction coding is that branch of information theory that concerns itself with getting the most bang for the bit, that is, with inventing practical, real-world codes that are as close as possible to perfect. Many error-correcting codes have been developed. They are used in virtually all consumer electronics devices that transmit, code, or decode digital information: digital telephone links, DVDs, audio CDs, computer hard drives, CD-ROMs, and more.

Sometimes the error rate that needs to be dealt with is low to begin with. For example, the industry standard for computer hard drives before error correction is one error for every billion bits read to or from the spinning magnetic discs inside the drive. This is a very quiet (low-noise) channel, but still too noisy for a computer. Modern computers read and write many billions of bits to and from their hard drives, so one error per billion bits might result in scrambled documents, crashed programs, messed-up money transfers, and e-mail sent to wrong addresses.

To prevent this, all computer hard drives use error-correcting schemes belonging to a family of codes called the Reed-Solomon codes. Reed-Solomon codes (named after their two inventors, who published the idea in 1960) are nearly "perfect" in the sense that they give almost the maximum amount of error correction possible for the number of redundant bits they add. When using an error-correcting code you cannot fit as much data onto a hard drive because of the redundant bits inserted by the code, but reading and writing from the drive suffers very *very* few errors.

Various versions of the Reed-Solomon code are used not only for computer hard drives but for audio CDs, DVDs, digital videotape, digital cable TV, digital cameras, and virtually every other commonplace digital data storage-and-retrieval device.

The most heroic deed of coding-theory history so far was performed by the Voyager spacecraft. Launched by the United States in 1977, *Voyager 1* traveled to Jupiter and Saturn and *Voyager 2* sailed past Jupiter, Saturn, Uranus, and Neptune. The two probes took sharp color pictures of scores of mysterious moons that are nothing but fuzzy points of light as seen from the Earth, even through powerful telescopes.

The Voyagers sent their pictures and other data back to the Earth as strings of ones and zeros. But they didn't have much power to do it with, so by the time a Voyager's signal arrived on Earth it was extremely faint. The largest radio antennas on Earth could only gather about 10^{-16} watts of power from a *Voyager 2* signal originating near Neptune, 4.6 hours away at the speed of light. This much power, if collected for three billion years, would light a 40-watt light bulb for less than one second. A digital watch uses billions of times as much power. Yet these ghostly signals from Voyager signals were detected, and color images of far-distant worlds were reconstructed from them.

This feat would have been impossible without error correction. For the Jupiter-Saturn leg of the journey, the Voyagers used a sophisticated error-correcting code called a Golay (24,12,8) self-dual code; *Voyager 2*, for its 1986 flyby of Uranus, was re-programmed to use an even more complex code, actually one code wrapped up inside another. The "outer" code was a Reed-Solomon code related to those used in CD players and other consumer electronics. Each of *Voyager 2*'s code words contains 32 bits, of which 26 bits are redundant bits. That is, over 80% of each Voyager code word consists of error-correction bits. These redundant bits enable computers on Earth to fix single and burst errors and even to replace "erasures" or completely lost bits.

The Voyagers are still transmitting data from deep space, leaving the solar system farther and farther behind with each passing second. And their messages are still surviving the trip thanks to error-correction codes designed on Earth using the principles of information theory.

Potential Applications

QUANTUM COMPUTING

Ordinary information is encoded in objects or signals that behave in a more or less reliable way. When a laser beam burns a microscopic pit on a music CD,

recording a single bit of information, that pit stays put. At least, it stays put until something physical, say a steel fork wielded by a younger sibling, comes along and changes it by force. Furthermore, each pit only means one thing, zero or a one, never both at once or some third, undefined thing.

That's ordinary information. There is also a thing called "quantum" information. Even computer scientists call it "weird" and are, according to the journal *Science*, "increasingly confused about how it works" (Sep. 14, 2001, Vol. 293, p. 2026). But the basic idea is not hard to understand.

Quantum information is still information, but it is called "quantum" because it is stored not by microscopic pits in a CD, or by voltages, magnetic fields, ink on paper, or any other physical object that has a definite state (there or not there, on or off, etc.). Instead, it is stored by objects like individual atoms, electrons, or photons. Such small objects, instead of obeying the physical laws of our everyday, human-scale world, obey the laws of the kind of physics called "quantum mechanics," which deals with the properties of atoms or smaller objects. The properties of such small objects are, by the standards of our everyday experience, simply crazy. For example, a basketball cannot spin to the left and the right at the same time, but an electron, according to quantum mechanics, can. If you can't see this in your mind, don't worry, physicists can't either. But they describe it mathematically, and the math always works, so we take it as a fact.

If, then, we use spin to store bits (say, leftward spin to store a 1 and rightward spin to store a 0), we can store a 1 and a 0 at the *same time* in the *same electron* (or atom). Physicists call a bit stored in this way a "qubit" (pronounced KYOO-bit), short for "quantum bit."

One reason computer scientists get excited when they think of qubits is that you can, because of superposition of states (the ability to spin left and right at the same time), run a single computation on quantum information and get twice as many answers as if it had been run on a conventional computer. And that's only the beginning. Groups of cubits can be linked to each other at a distance by the quantum effect that physicists call "entanglement." With a boost from entanglement, N qubits can do the computational work not just of $2N$ but of 2^N conventional bits. That makes computation with qubits not just twice as fast as conventional computation, but *exponentially* faster.

But there is a good reason why we don't all have qubit-based PCs and Macs on our desks, and that is that it is not easy to teach an individual atom to sit up, roll over, and bark ones and zeroes on command. Researchers are approaching the problem by using electromagnetic fields to bottle up small numbers of ultracold atoms, then zapping them gently with laser beams, a procedure that does not involve not the kind of hardware you easily can scoot out and buy at the local electronics store. Further, qubits tend to "leak" or evaporate, losing their information. So ticklish is quantum computation that although it has been studied for over 30 years, the first demonstration of a true quantum computation was not made until 2002, and even that only involved a few bits. Progress remains slow but the potential payoffs are great, so research continues. Computation using quantum information teases us with the possibility that computers may someday billions or even trillions of times faster than today's.

Where to Learn More

Periodicals

Bennet, Charles H., and David P. DiVincenzo. "Quantum Information and Computation," *Nature*, 16 March 2000, Vol. 404, pp. 247–255.

Seife, Charles. "The Quandary of Quantum Information," *Science*, 14 September 2001, Vol. 293, pp. 2026-2027.

Shannon, C.E. "A Mathematical Theory of Communication," *The Bell System Technical Journal*, Vol. 27, pp. 379-423, 623-656, July, October, 1948. Available online at <http://cm.bell-labs.com/cm/ms/what/shannonday/shannon1948.pdf.>

Websites

Touretzky, David S. "Basics of Information Theory." Computer Science Department, Carnegie Mellon University <http://www-2.cs.cmu.edu/~dst/Tutorials/Info-Theory/> (September 28, 2004).

Inverse

Overview

If something is transformed by a sequence of events, the inverse is another sequence of events that will bring it back to the start. The nature of the inverse steps, compared to the steps that are applied, allows identification of how the system is related to its starting conditions under the transformation. The definition of the inverse is of fundamental importance and is probably one of the first mathematical entities that would be sought in many problems. The nature of the inverse allows one to understand how the transformation will act when trying to manipulate it. This concept is important in understanding the algebra that can be applied to the transformation in a mathematical analysis. The nature of the inverse has implications and applications that span all areas of mathematics, science, and engineering, and will often be defined in many different forms.

Fundamental Mathematical Concepts and Terms

DEFINITION OF AN INVERSE

If a sequence of actions is applied, then by definition, the inverse is the sequence of events that undo these effects, returning to the start. For example, imagine a length of rope, the sequence of events that will be applied include tying a knot in the rope. The inverse of this operation will then be the sequence of events that untie the knot and leave the rope as it was originally. This is where the term inverse comes from, as we invert what has been done. If R is the series of actions that were applied to something, F, then the inverse sequence is written by adding a superscript: $R^{-1}(RF) = F$.

THE MULTIPLICATIVE INVERSE

It is now possible to investigate some interesting forms of the inverse. Consider if the reverse of the above is true. In this case, we swap the order of our actions by applying the inverse first, followed by the original sequence of events: $R\,(R^{-1}F)$. If the result is also identical when nothing is done, the events are their own inverse. They are said to be a multiplicative inverse, or that they commute. For example, consider rotating a photograph to the right through some angle and then its inverse, to the left, so that it is back to where it started. Now, swap the order of the events by applying the inverse, rotating the photograph to the left this time, followed by the original rotation to the right. This will again leave the photograph unchanged. This is written as $R^{-1}(RF) = R(R^{-1}F) = F$.

OPERATIONS WITH MORE THAN ONE INVERSE

Now consider tossing a coin. If the coin always starts as heads before it is flipped, there are two possible outcomes and hence, two inverses. Approximately fifty percent of the time, the inverse action would be to do nothing, as the coin lands on heads. For the other fifty percent of the time, the inverse action would be to turn the coin over after it lands on tails. However, it is not possible to know which inverse to choose before the coin is flipped, as at this point, both outcomes are just as likely. In this case, the multiplicative inverse does not hold true. The multiplicative inverse only holds for actions that have one possible inverse.

OPERATIONS WHERE THE INVERSE DOES NOT EXIST

Sometimes the steps needed to find the inverse may be so complicated that we can assume that it does not exist. An example involves hitting a pack of balls on a snooker table forcefully. After hitting the pack, the balls will be spread all over the table. The inverse of this would be to give all the balls a shove in the opposite direction so that they all roll back into a pack, just as if a film of the shot were run in reverse. However, if we tried to do this in reality, small errors in the velocity and imperfections on the surface of the balls and table would be magnified as they moved in reverse, with the result that no matter how hard you tried, the balls would never form the original pack shape again. In this case, the inverse will be impossible to perform in reality.

INVERSE FUNCTIONS

Consider a set of points along the x axis, 3, 5 and 9. The function $y = f(x) = 3x + 2$ will transform these points to the following coordinates *(x,y)*: (3,11), (5,17), and (9,29).

An inverse function would have the effect of returning each y axis value to the original x axis value. Looking at the formula, it is seen that this is achieved if it is rearranged by adding 2 to both sides and dividing by 3, as in $x = f^{-1}(y) = (y - 2) / 3$.

If the function and its inverse function are substituted for each other, they are seen to leave the coordinates unchanged as expected, $f^{-1} f(x) = (3(x + 2) - 2) / 3 = x$.

If a graph of the function is made, the effect of the inverse function will be to reflect all the points, *(x,y)* along the line $y = x$ (see Figure 1).

Now consider the function $f(x) = x^2 + c$.

By definition, a function can only have one result and functions such as $y = x^2$ are said to not have an inverse. It can also be seen that $y = x^2$ will not satisfy the multiplicative inverse relation. This is not quite as bad as it sounds because in real life applications, boundary conditions can be put on the range of values a function can take. In this case, we may find that in our analysis x can only take positive values; maybe it is a radius or mass where negative values would make no sense. It is then safe to ignore the negative square roots and define an inverse within these boundaries.

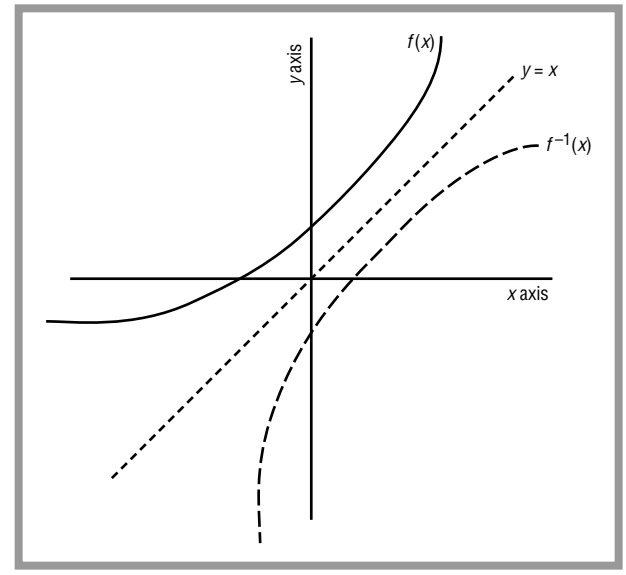


Figure 1: Reflection along $y = x$ of the inverse function.

A Brief History of Discovery and Development

An inverse operation is such a fundamental idea that examples are found all through the history of mathematics, geometry, and science. One of the more historically interesting uses is in code breaking. All throughout history, codes have been invented and deciphered in attempts to gain some form of political or military advantage. In the time of the French king Henry the IV (1553–1610), Spain had ciphers that they changed regularly and was thought impossible to break. Henry IV gave the problem to a mathematician called Vieta, who figured out how to decode it and it was used by the French for two years to read Spanish documents. It was so successful that King Philip II of Spain (1527–1598) complained to the pope that the French were using sorcery to read their communications.

Another interesting fact is the development of negative numbers, used as the inverse of addition. It would

Inverse Square Law

The inverse square law is used to describe a field generated by a point charge, where the field is free to extend in every direction into free space. The field generated by a uniform sphere will be identical to the point charge field and at large enough distances; non-uniform shapes will often become good approximations to a point source as well. This means that the inverse square law accurately describes or gives very good approximations to many fields generated in nature. Examples are the gravitational fields generated by stars and planets, sources of radiation such as light and heat, and the electromagnetic forces such as those between atoms.

At a given distance, *r*, from a source the inverse square law is given as the intensity of the source *A*, divided by the area of a sphere,

$$I = \frac{A}{4\pi r^2}$$

For example, if one moved two feet away from a spherical light source, the intensity of the light would drop by a factor of four. To invert this, and keep the light intensity the same as before, the intensity of the light source would need to be increased by a factor of four. This explains the reason why a lighthouse focuses the light into a beam of light. This is not a point source, as the light has been kept in a narrow beam, and the intensity will only fall off as the light beam spreads due to focusing imperfections or atmospheric effects and will be visible to much greater distances.

seem that such a common concept would have been used since ancient times, but in fact it was not until 1545 that they came into common use. It was Gerolano Cardano (1501–1576), an Italian mathematician, who first showed that negative numbers could be used as an extension of our number system and that this was useful in the calculation of debts for example.

Real-life Applications

CRYPTOGRAPHY

Cryptography is the science of encoding information so that it can be transmitted secretly without an eavesdropper decoding it. Consider a simple cipher when encoding a letter to a friend by swapping all the letters and spaces by numbers. The friend has a similar sheet of numbers to letters that they use to decode your message. This sheet is called a key. In this case, the inverse steps are equal to the original number of steps you used to encode the message. However, an eavesdropper can decode your message by hand or writing a computer program that randomly assigns letters to numbers and scans for words in a dictionary. This is repeated until parts of a word is found that you can guess, and as more words are found, it will become possible for the eavesdropper to regenerate the key and decode your message. It is possible to use more complex functions to encode your message, but it can be shown that an eavesdropper will always be able to find the inverse function in a length of time that gets shorter the more the transmitted text becomes larger than the key. By sophisticated use of mathematics and computers, it is obvious this is not the most secure method to transmit data, and that it becomes weaker the more it is used.

A method called public key encryption is very common nowadays, and can get around this problem wherever information needs to be sent securely over unsecured lines. For example, Internet banking and automatic teller bank machines use public key encryption. Public key encryption works as follows. Imagine Bob wants to receive information from Alice. Bob has a special function that is split into two parts, the public function and private function, and generates keys using these functions. Bob sends the public key to Alice over an ordinary unsecured line and keeps the private key. Alice encodes the information she wants to send to Bob with the key using her public function and sends the resulting sequence of code back to Bob over the unsecured line. This public function has no inverse, so even if an eavesdropper were to gain access to the function, the public key alone provided no information.

When Bob receives the encoded data from Alice, he combines the private and public keys, to generate the inverse function. By running the information through this new function, the message from Alice is recovered. If Alice wants to receive information from Bob, then the reverse scenario is used and a completely secure two-way conversation has been set up over public lines. The reason this system works is that there is no analytical way to generate the inverse from the public key, but it can be generated with the combination of public and private keys. The only way it is thought possible to break this encryption would be to use the theoretical quantum computer. Until such a device exists, this form of encryption will remain theoretically impossible to break.

Multiplicative Inverse

As described previously, the multiplicative inverse is satisfied by actions that only have one possible inverse as $R^{-1}(RF) = R(R^{-1}F) = F$.

An obvious example of such a function is multiplication, which has the inverse of division. Consider multiplication by the number 3, which has the inverse ⅓ as in $(\frac{1}{3}.3).F = (3.\frac{1}{3}).F = F$. Multiplication by ⅓ is really division by 3—the inverse of multiplication by 3.

Another example is addition. The inverse of this is subtraction. Consider the action of adding 3. Here $R = +3$ and the inverse R^{-1} is subtracting 3: $(F + 3) - 3 = (F - 3) + 3 = F$.

Even though this system was known about for many years, it was only the invention of such functions by the English mathematician Clifford Cocks in the 1970s that made it possible. It was promptly made a state secret until it was independently invented by several American mathematicians in 1976. The American government made the code military property and legal battles ensued over about the encryption method's future. However, control of the encryption method was finally defeated in 1991 after versions of the code for Pretty Good Privacy (PGP) were published. In some countries, this method of encoding information is still illegal.

NEGATIVES USED IN PHOTOGRAPHY

When a black and white photograph is taken with a camera, light falling on the film will turn the film black, and areas where no light falls will remain clear. After the photograph has been taken, the film is chemically fixed so that it no longer responds to light and the image has been recorded. However, on examining the film, the image will be the inverse of what is desired; dark areas will be light and light areas dark. In this case, the photographic film is called a negative. In order to produce the final image, the areas of light and dark must be reversed. This is done with a special camera that will pass light through the negative and onto photographic sensitive paper. By retaking the image again in this way, the areas of light and dark are inverted and the original image is returned. The second step of re-exposing the negative of the image is equivalent to taking the inverse. Although technically more complicated, the process is the same for producing a color image.

THE BRAIN AND THE INVERTED IMAGE ON THE EYE

The eye works by passing light through a single lens and focusing it on the retina at the back of the eye. Due to the nature of optics, this image is upside down. The reason humans do not see the world upside down is that the brain inverts the image, therefore, the world is seen right-side up. It was thought that this process was hard-wired into the human brain for many years until a series of experiments were conducted that suggested otherwise. In these experiments, subjects wore a special mask over their eyes for 24 hours a day for several weeks. This mask placed lenses in front of the subjects' eyes that caused them to see the world upside down. For a period of time, the subjects were naturally disorientated and confused, but after a while, started to see the world the correct way again, suggesting that the human brain had the ability to re-configure itself to cancel the effect of the mask.

FLUID MECHANICS AND NONLINEAR DESIGN

In the snooker ball example discussed previously, it was seen that after hitting a pack of snooker balls, there existed no inverse operation where we could give the balls a shove and they would roll back into the formation of a pack as any tiny error would be magnified and ensure that the balls were always distributed randomly over the table. This has an important consequences for industrial design.

Consider using a computer to optimize the flow of fuel from a nozzle for an engine you are trying to design. At low velocities the flow from the nozzle will be smooth, called laminar flow and the computer simulation will accurately reproduce the flow. However, the laws of fluid dynamics are non-linear, i.e, they do not have a simple inverse and the flow at future moments is not related to the flow to past moments in a simple way. As a consequence any errors in the measurements will start to multiply rapidly and at a certain flow rate this is seen as a turbulent flow where any predictions from the computer will rapidly deviate from the real situation. In this case, it is necessary to find a specific approximating model to the nozzle that you are trying to design.

In industry, there are many examples of this situation, such as modeling the flow and timing of inks from printer heads, gas flow in exhaust systems to the effects of rain and dirt on glass. Even though the physical laws of each situation may be written down in a few lines, the non-linearity means that special and often highly technical models have to be constructed to investigate each situation and this can involve much time and investment making the details of such models closely guarded industrial secrets.

ANTI-SOUND

Sound is a series of pressure waves in air. Sound that is not loud enough to cause damage to the hearing can still cause discomfort and irritation after long periods of exposure. In modern design keeping sound down to minimal levels is a key concern in commercial passenger transport, such as ships, trains, planes and cars. By reducing the noise most passengers will experience a more present and relaxing journey. Another area where sound levels need to be kept under control is in military equipment where the high performance needed will often result in very noisy equipment but large amounts of sound absorbent material is undesirable due to weight and space concerns.

One solution to reduce sound in these environments is to generate anti-sound. This involves wearing earphones or fitting the passenger compartment with loudspeakers that generate a special form of sound. Speakers will record the sound waves in the compartment and a computer will calculate the exact pressure waves that will cancel these pressure waves, called inverse or anti-sound. When the anti-sound plays through the speakers it will reduce the sound that reaches the ear of the passenger. Although conceptually simple, in practice these systems are very difficult to make, for example real time anti-sound generation requires rapid computing and playback times, and can be very expensive and sensitive to changes in its environment. Some systems get around this by using a digital recording to playback the anti-sound. In practice, it will be impossible to completely cancel all the background sound, but the systems have been shown to be very effective at reducing background noise levels.

STEALTH SUBMARINE COMMUNICATIONS

Communicating with submarines is not as easy as communicating with space craft. Unlike space and air, water is a much denser medium. In pointing radio signals at submarines, they will be rapidly absorbed by the water. Another problem with submarine communications is that one of their main strengths comes from being hidden. For example, submarines make up a key part in the nuclear defense systems of both the East and West, and can remain hidden off a coastline or under an ice flow for many months at a time, if needed, before engaging their target. An exposed submarine is vulnerable to missiles from other submarines, surface ships, and aircraft. Submarines are also vulnerable to attack, as explosions under water can be far more devastating even if there is not a direct hit due to the large pressures on the hull. For this reason, high-power focused radio transmissions are a problem, as the direction of the beam can give the position of the submarine away to the enemy.

One system for communicating with submarines relies on a series of radio antennas that transmit at extremely low frequencies (ELF). It is a physical property of radio waves that their absorption by water increases with frequency. By using ELF transmissions at around 76 Hertz, the radio signal can penetrate the water to depths of hundred of feet as opposed to a similar VHF transmissions would only penetrate to a depths of around 10 feet (3 m).

When setting up a system like this, there are multiple transmitters to give a wide coverage. Having multiple transmitters of such low frequency can cause blank spots in radio transmissions in areas where a signal would be expected. This is because the waves from each transmitter interfere with each other, adding to the overall power at one point or diminishing the overall power at another, forming an interference pattern. This can be thought of as throwing two stones into a small pond. After the initial splash, the ripples will form a stationary pattern on the water's surface, with some areas having higher ripples and some areas that look calmer. The interference pattern generated with such low-frequency transmissions will leave blank spots in a signal that can be larger then the submarine, and a careful design of the antennas is needed to give constant communications with the submarine.

Blank spots in the reception can be shifted if the timing of the transmitted signals, called the phase of the transmission, is changed in one or more of the transmitters. Now imagine if the submarine is in a certain position in the sea and it transmitted a signal. This signal would spread out and reach all of the transmitters slightly different times, i.e., with different phases. If a signal is transmitted back to the submarine as the inverse of the way a signal is received, then the signals from the transmitters can be made to interfere with each other and build up as they approach the submarines location. This effect that can be thought of as the inverse of ripples formed from throwing a stone into a pond, in this case, the radio ripples move toward the center of the pond, building up into the original splash, at the location of the submarine.

As the submarine moves, the phase of the transmitters can be changed to refocus the transmission to the new location. At other points on, the radio signals will be more difficult to detect, making the system more secure against eavesdropping.

STEREO

Having two ears, the human brain can figure out the location of a sound by the time delay in the signal

as it reaches either ear. For example, if a sound reaches the left ear first and then the right ear a fraction of a second later, the brain interprets the source closer to the left side and this is the feeling we would have from this source.

To record a sound in stereo, two microphones are needed and these are placed side-by-side at about the distance of the human head in front of the artist. Having the two microphones means that not only is the sound recorded, but the delay phase of the sound is also recorded. In a modern studio, equipment exists that can add this delay electronically. In this manner, artists can record themselves with high-quality microphones and pickups onto separate channels, and the studio can electronically add phases to each channel and make the artists sound as if they had been playing at different positions in a group.

To play back sound in stereo, two speakers are placed in front of the listener to invert the effect of the recording. During playback, the timing or phase of the sound waves from either speaker will interfere with each other, some waves adding to each other (called constructive interference) and some waves canceling each other (called destructive interference). The interference of the waves as they reach the ear will result in time delays in the sound between each ear, which the brain will reconstruct as positions of sound sources. It is possible to hear the musicians playing from different positions in the group as if they were in the room with the listener. For this system to work properly, the locations of the speakers and listeners, and reducing sound reflections from the walls of the room will all be crucial factors in how well the stereo effect is reconstructed. It can take much time and investment to construct the ideal listening environment, especially in environments such as cinemas, where there are many listeners to consider.

Where to Learn More

Web sites

Fischer, Charlie. "Public Key Encryption." <http://www.krellinst.org/UCES/archive/modules/charlie/pke/> (April 9, 2005).

United States Navy. "Extremely Low Frequency Transmitter Site, Clam Lake, Wisconsin." *Navy Fact File* <http://enhttp://www.elfrad.com/clam.htm> (April 9,2005).

Wicks, J. "Details about the functional inverse." North Park University. <http://campus.northpark.edu/math/PreCalculus/Functions/Functions/Algebra/> (April 5, 2005).

Iteration

Overview

Iteration is a process of calculation that is repeated again and again, each time improving the accuracy of the result by using the output of each step for the input of the next iteration. By repeating the process and analyzing successes and errors, humans and even some machines can improve at performing the action.

Fundamental Mathematical Concepts and Terms

An example of problem solving by the process of iteration could be found in the following set of computer programming instructions: "While a number is less than 15, continue printing that number but increase its value by one. Upon arriving at the number 15, activate the programmed screensaver." These instructions could well relate to the set up of a screensaver to come on after a computer screen was been inactive for 15 minutes.

Let's look at the nuts and bolts of the above set of instructions in a bit more detail. If the starting number was two, for example, it would be less than 15, and so the statement would be judged as being true (i.e. less than 15). That would be a signal to go through the process again, and again, until the number was 14. At that point the number plus one would no longer be less than 15. That would be the signal for the computer to switch to another set of instructions that activates the screen saver.

Real-life Applications

ITERATION AND SPORTS

You do not have to be a big basketball fan, or even have much of a knowledge of the game, to recognize the name Michael Jordan. Michael Jordan was an awe-inspiring basketball player. One big reason was talent, but another was iteration. A particularly relevant example of this process was Jordan's routine at the end of each practice. Jordan would shoot the basketball at the hoop from various locations, near to the basket and further away. His practice was not over until he sank 100 shots in a row. If he made 92 baskets and missed one, he would start over again. This routine, which must have been frustrating on a less-than-accurate day, is analogous to mathematical iteration. By performing the action over and over again, using the results of each shot to refine the performance of the next shot, Jordan fine-tuned a technique that predictably put the ball through the hoop.

Similarly, Tiger Woods has used thousands of hours of golf practice to perfect a golf swing that is consistent from one day to the next. This consistency propelled him to become the number one ranked golfer in the world in 2003.

But even the best golfer in the world likes to tinker with his swing, to try out slightly different changes in hopes of producing a swing that is even better than before. That is the essence of iteration. By repeating an action again and again, changes can be evaluated and, if they are successful, can be incorporated into the action.

ITERATION AND BUSINESS

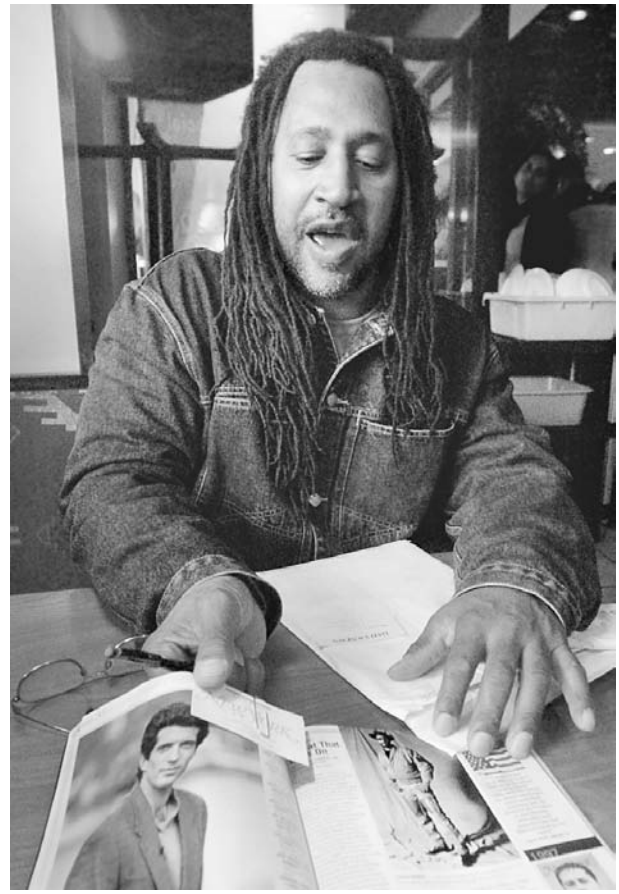
Not surprisingly, iteration is a favorite buzzword of computer programmers. Computer programmers often make available trial software programs on the Internet, called beta versions. Beta programs are a form of practice versions of a new program. Usually this iteration of a program has more features built in that presumably will make people want to buy it and use it instead of the current version of the program. The purpose of a beta version program is to encourage people to try the software, figure out its good points and, most importantly, discover what needs changing or what does not work. The software programmers can then change the beta version to produce the final improved program that is widely sold.

In another business application, iteration is an important feature of accomplishing a project that involves a large team of people. Again, in the realm of computing, an ideal example of iteration involves extreme programming, or XP. Like an extreme sport, XP is a difficult-to-accomplish form of programming that often involves dozens of programmers. These programs are updated frequently and made available much more often than, for example, a program for a video game that might be updated once every three or four years.

In the XP iteration process, the total project is usually broken into chunks. Each chunk can have a back-and-forth process where the component of the program is written, tested, and returned for tinkering. A tight schedule allows the iteration process for each chunk of the project to be accomplished by a deadline, so that all the chunks can be put together to produce the final product. As well, the back-and-forth contact between people that is part of this type of iteration allows for better tracking of minute details in the frenzy of the project.

ITERATION AND CREATIVITY

Creativity involves the ability to look at something in a different way, to find a new idea. A necessary part of creative thinking is gathering information, and then trying to put that information together in a new way. This is where iteration comes in. New product ideas come to the forefront after cycles of inventing a design, testing the design, and, as usually happens, discovering and fixing problems.



DJ Kool Herc, the Jamaican-born DJ considered the father of hip-hop. It was Herc, at parties in the early 1970s, who began playing the instrumental segments of songs over and over again, a form of musical iteration, while speaking in rhyme over them. AP/WIDE WORLD PHOTOS. REPRODUCED BY PERMISSION.

Iteration in creative product design can be illustrated by considering a new CD from a popular music group. The tracks on the CD do not usually happen in one recording session in the studio. When the band first starts to record a song, the musicians, songwriters, and producer may have different ideas of what the final version will sound like. Different versions (iterations) of the song are tried out, discussed, and changed until the artist is pleased with the final version. The final track that is heard on the CD is often very different from what a band member thought it sound like months before.

People who are known for their creative approach to their work often say that the process they use to come up with all those great ideas is very structured. They take the same approach to each problem, knowing that doing the

steps in much the same order (there has to some flexibility in how things are done) helps their mind get ready to think. In other words, their whole approach to being creative involves iteration. Similar actions are repeated.

ITERATION AND COMPUTERS

In a computer program, iteration is the recycling of a set of instructions, known as looping. A single iteration is one pass of the instructions. Once the set of instructions has been written, a computer will quickly pass through the loop over and over again without making mistakes (unlike humans).

Another real-world example of iteration in computing is a macro. A macro is the putting together of a series of commands that responds to one signal (like the pressing of a designated key on the keyboard). Macros are not necessary to do work on a computer, but they make time go more quickly. Instead of typing in the same commands over and over, this iteration is taken care of by the one action of pressing that designated key.

Iteration and Nature

In nature, repeating patterns are common. From the spirals of a seashell to the icy beauty of a snowflake, to the many hexagons of the honeycomb of a beehive, a basic unit is repeated again and again to produce the final structure. This repeat of the basic unit is iteration.

A cutting-edge example of iteration in the laboratory is molecular cloning. Molecular cloning involves creating a genetic twin from the genetic material obtained from a living creature. Experiments in plants and animals have shown that scientists have not yet perfected the clon ing process. When they do, then cloning will be a living example of iteration. Whether this form of iteration is desirable or not is being debated at the present time.

Where to Learn More

Books

Woods, T. *How I Play Golf.* New York: Warner Books, 2001.

Web sites

Peterson, Ivars. "Ivars Peterson's Math Trek: Candy for Everyone" *Mathematical Association of America* <http://www.maa.org/mathland/mathtrek_1_10_00.html> (September 5, 2004).

Wells, D. "Iteration Planning." *Extreme programming* <http: //www.extremeprogramming.org/rules/iterationplanning.html> (September 3, 2004).

Linear Mathematics

Overview

Linear mathematics deals with linear equations. An equation is "linear" if it consists of a sum of variables or unknowns, each of which is multiplied by some number or constant (examples will be given below). Many real-world problems in physics, engineering, business, chemistry, biology, and other fields are described by linear equations. Computers are use to solve linear equations in groups or "systems," making possible many kinds of medical and scientific imaging, realistic video games, cheaper design of cars and other products, and the more efficient management of money.

Fundamental Mathematical Concepts and Terms

Linear equations are called "linear" (line-like) because the simplest kind of linear equation—one having two variables—describes a straight line. For example, the equation $2x_0 + 3x_1 = 4$ describes the straight line depicted in Figure 1.

Here x_0 and x_1 are "variables," meaning that they stand for any numbers we like; the small 0 and 1 are labels to tell them apart by. For each x_0 we choose, there is one and only one x_1 that makes $2x_0 + 3x_1 = 4$ true. For example, if we set x_0 equal to 0, then x_1 must be 4/3 because:

$$\underset{\substack{\uparrow \\ x_0 \text{ value}}}{2\times 0}+3\times\underset{\substack{\uparrow \\ x_1 \text{ value}}}{\frac{4}{3}}=4$$

We can also use letters to stand for the fixed numbers that multiply x_0 and x_1. If we replace 2 and 3 in $2x_0 + 3x_1 = 4$ with the symbols a_1 and a_2, and replace 4 with b—where these new letters can stand for any fixed numbers we like—we get a general-purpose linear equation in two variables: $a_1x_1 + a_2x_2 = b$.

We can extend this to as many multipliers (also known as "coefficients") and variables as we like. The equation is still called a "linear" equation no matter how many variables we add. Here is the form of linear equation involving 3 variables and 3 coefficients:

$$a_1x_1 + a_2x_2 + a_3x_3 = b.$$

We have already seen how a two-variable linear equation describes a line. A three-variable linear equation

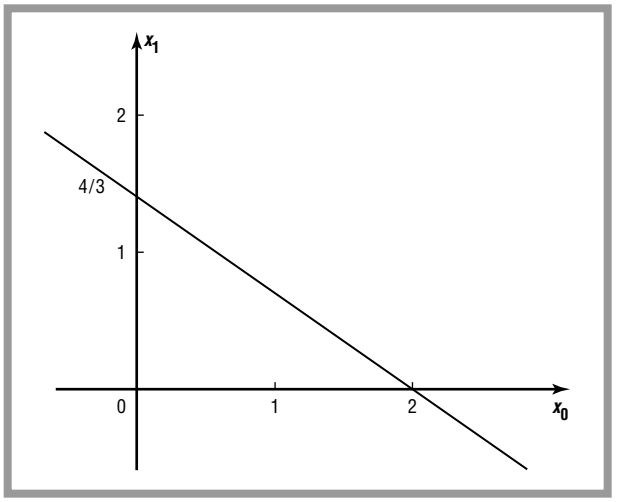


Figure 1: Graph of the equation $2x_0 + 3x_1 = 4$.

describes a plane, a set of points resembling a stiff sheet of paper tilted in space.

In general, a linear equation containing *n* variables and *n* coefficients looks like this:

$a_1x_1 + a_2x_2 + a_3x_3 + a_4x_4 + \ldots + a_nx_n = b.$

The three dots in the middle of the equation stand for all the terms between the fourth term and the *n*th term that we don't want to bother to write down. In real-world applications, linear equations containing dozens or even millions of terms are common.

Linear equations can be combined into groups or systems. A system of linear equations is a group of two or more equations that involve the same variables. The following is a system of two linear equations involving the two variables, x_0 and x_1:

$$2x_0 + 3x_1 = 4$$
$$x_0 + 9x_1 = 0$$

The "solution" of a system of linear equations is that set of numbers which, if plugged in for the variables, makes every equation in the system true at the same time. In this example, the solution of the system is $x_0 = 12/5$, $x_1 = -4/15$. This solution is unique; that is, each equation considered by itself is true for many values of x_0 and x_1, but only at $x_0 = 12/5$, $x_1 = -4/15$ are both equations true.

If you graph the two equations in this system as lines on paper, the solution of the system will be the point where the two lines intersect. Every system of equations has a single, unique solution (like this system), or no solutions, or an infinite number of solutions. Among systems that consist of two lines, like the ones that we've just been looking at, those that have a single, unique solution are lines that intersect (one point in common); those that have no solutions consist of parallel lines (no points in common); and those with an infinite number of solutions consist of two equations for the exact same line (all points in common).

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

Figure 2: A "two-by-two" matrix.

Systems of equations can also be written as matrix equations. A matrix is a rectangular array of numbers or variables with square brackets around it. It is named according to how high and how wide it is. For example, the matrix shown in Figure 2 is a 2×2 ("two by two") matrix because it is 2 entries tall and 2 entries wide. The matrix depicted on Figure 3 is a 2×3 ("two by three") matrix because it is 3 entries tall and 2 entries wide. A matrix can be added to, subtracted from, or multiplied by other matrices. It can also be multiplied by numbers, variables, and vectors, which are special matrices only 1 entry wide. Vectors containing three entries, as depicted in Figure 4, are particularly useful in science, engineering, and computer animation because each three-entry vector can specify a point, force, velocity, or acceleration in three-dimensional space.

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix}$$

Figure 3: A "two-by-three" matrix.

$$\begin{bmatrix} 3 \\ 2 \\ 1 \end{bmatrix}$$

Figure 4: Vector containing three entries.

The system of two equations shown earlier can be written as a 2×2 matrix multiplied by a vector and set equal to a second vector. That is,

$$\begin{aligned} 2x_0 + 3x_1 &= 4 \\ x_0 + 9x_1 &= 0 \end{aligned} \quad \text{means the same as} \quad \begin{bmatrix} 2 & 3 \\ 1 & 9 \end{bmatrix} \begin{bmatrix} x_0 \\ x_1 \end{bmatrix} = \begin{bmatrix} 4 \\ 0 \end{bmatrix}.$$

When there are only two or three variables in a system of equations, as in this example, there is no advantage in using the matrix form. But when systems involve many variables, as in most real-life applications, the matrix form is more efficient and revealing. Computers are well-suited to calculating with matrices, and are often used to solve systems whose matrices contain millions of entries. In creating medical images of the inside of the body, searching for oil reserves, predicting global climate change, designing new drug molecules, maximizing profits, and many other applications, the solution of large matrix equations by computer is key.

Because the solution of systems of linear equations is so important in our high-technology society, most of the examples of linear math given below involve the solution of such systems.

Linear equations

Linear equations that involve two variables, such as $2x + 3y = 4$, describe straight lines. That is, if you graph any of the x, y pairs that satisfy the equation, you will find that they all lie on the same line on the paper—and, likewise, that every point on that line satisfies the equation. A linear equation that involves not two but three variables, such as $2x + 3y + 7z = 4$, graphs a plane in three-dimensional space.

Linear equations appear everywhere in science, technology, and business. If you are selling sneakers at x dollars of profit a pair, you know that if you sell 20 pairs of sneakers you will make double the money than if you sell 10 pairs, namely, $20x$ dollars rather than $10x$ dollars. Here the relationship of pairs sold to total profit is described by a linear equation: number of pairs sold (call it a) times profit per pair (x) equals total profit (p), $ax = p$. Anyone running a lemonade stand knows this much linear math by instinct.

But not everything in real life is linear. For example, you might make more profit per pair of sneakers if you sell a million pairs than if you sell only a hundred. In this case, the equation describing your total profit in terms of sales will not be a linear equation.

Nor are all linear equations as bare-bones as $ax = p$. If you are selling two types of sneaker, one of which makes x_1 dollars of profit per pair while the other makes x_2 dollars, a different linear equation arises. Say you sell a_1 pairs of the first type of sneaker and a_2 of the second type. Then your total profit p is given by the sum of the profits from each type: $a_1x_1 + a_2x_2 = p$. This is also a linear equation, but it involves two variables. In real business and industry, equations of this type involving variables are common.

Real-life Applications

EARTHQUAKE PREDICTION

Science foresees no way of preventing earthquakes, which occur when whole sections of the Earth's crust, many miles across and weighing billions of tons, slip past east other. These forces are too great to control. However, knowing when and where earthquakes are likely to happen, and how strong they are going to be, would make better preparedness possible and reduce the loss in lives and money caused by major quakes.

It is not yet possible to predict most earthquakes, but with the help of large systems of linear equations solved by computers, scientists are making rapid progress. The basic method, called "finite element modeling," is a common one in manufacturing, science, medical imaging, and other fields today. In finite element modeling, a mathematical model or image of an object or volume of space is built up using either triangles (for flat models) or tetrahedra (four-pointed pyramids, for three-dimensional models). The triangles or tetrahedra are called "elements" and fit together into a web or network called a "mesh." One or more separate variables (like x_0, x_1, and so forth used above) are assigned to each element, and linear equations involving these variables are written so that they approximate the laws of physics that apply in that area of space. For earthquake prediction, meshes containing 100 million tetrahedra or more are created that represent parts of Earth's crust containing earthquake faults. Equations are constructed using these meshes that describe how shock waves move through the rock and soil. Supercomputers are then used to solve the resulting systems of millions of linear equations; the solution shows what an earthquake will look like.

Linear Inequalities

Equations express equalities, such as $1 + 2 = 3$. We can also write *in*equalities, expressions that say that one thing is less (or greater) than some other. Four signs are used to express inequality: $<$ (less than), $\leq$ (less than or equal to), $>$ (greater than), and $\geq$ (greater than or equal to). For example, the expression $a > b$ reads "a is greater than b." The expression $c \leq a$ reads "c is less than or equal to a."

A linear inequality is a linear equation with its equals sign replaced by an inequality sign. The linear equation, $x + y = 2$, for example, can become the linear inequality $x + y \leq 2$.

The linear equation $x + y = 2$ describes a straight line—and that's why it's "linear" (See Figure A.)

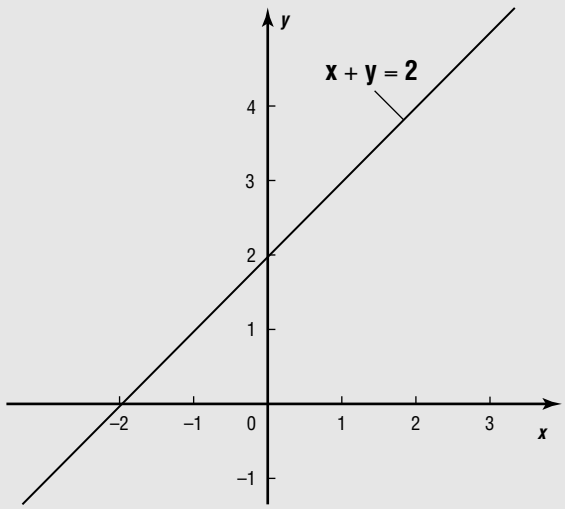


Figure A: Graph of $x + y = 2$.

The linear inequality $x + y \leq 2$ also describes a set of points. The line $x + y = 2$ is part of that set because the "less than or equal to" sign includes an equals sign.

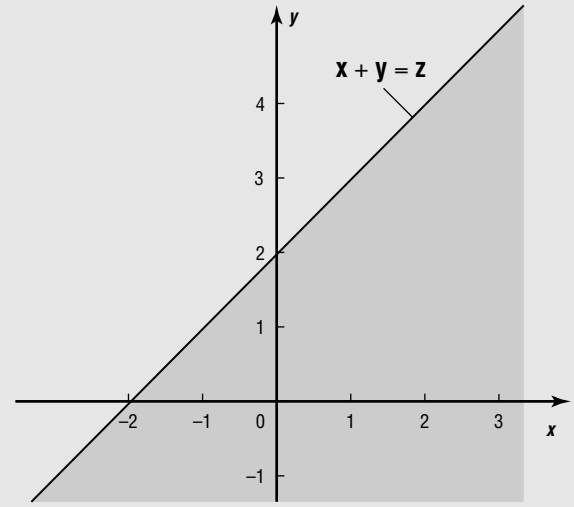


Figure B: Graph of x + y = 2.

The inequality is also true for all points below the line, namely the gray area in Figure B.

Linear inequalities arise often in real life. Consider a factory that can make two kinds of computer chip, Chip One and Chip Two, but cannot make both at the same time. The time needed to make a batch of Chip One is 5 minutes, so the time to make c_1 batches of Chip One is $5c_1$ minutes. The time needed to make a batch of Chip Two is 10 minutes, so the time to make c_2 batches of Chip Two is $10c_2$ minutes. But there are only 1,440 minutes in a day. Adding the time spent in one day making Chip One to the time spent making Chip Two, we have the linear inequality: $5c_1 + 10c_2 \leq 1{,}440$. This example is simple but not far-fetched. Linear inequalities that express limits or constraints on time, material, or other valuables appear constantly in the solution of real-world business and finance problems. Such problems are often solved using the technique called "linear programming."

RECOVERING HUMAN MOTION FROM VIDEO

There is much interest today in teaching computers how to track human motion from video cameras. To track human motion successfully, a computer must be able to pick human forms out of all the other information in a moving image, like the video of a dance or a football game. It should then be able to describe what the human being has done in words, or be able to move a mathematical model or virtual puppet to re-create the motion it has observed. The ability to track and then describe or reproduce human motion mathematically is used in video games, virtual reality, formation analysis in sports, and in other ways.

One method used to track motion is the identification of "feature points" on the subject—idealized dots or spots on the surface of the subject's body, say on their helmet or elbow or knee. These feature points are then tracked in the video images recorded by several video cameras. Since each video image is two-dimensional (flat), the location of each feature point in the image at any one time can be described by two numbers, an x coordinate that says how far from the left-hand edge of the image the point is and a y coordinate that says how high up it us from the bottom edge of the image. If there are, say, 40 feature points on the subject, there are then 80 numbers that describe where the feature points are at a particular moment of time in the video from a single camera, and $3 \times 80 = 240$ numbers to describe where the feature points in the video from 3 cameras. These numbers are put into a matrix. Using linear mathematics methods of matrix algebra, this matrix is separated into two matrices, an M matrix that describes how the cameras are pointing (which we don't really care about) and an S matrix that describes the true arrangement in space of the feature points. The S matrix records how the subject is positioned in space at that moment. A whole series of S matrices, one for each frame of video, describes how the subject's body moves through space over time. Motion capture and analysis using linear algebra is used in computer-animated movies such as *Polar Express* (2004), where live actors' motions were recorded by computers and then used to animate digital figures.

A check sorting machine separates checks at high speed at the Unisys Corp. check-processing facility. Banks use linear programming to process checks more efficiently. In particular, they want to minimize "float." AP/WIDE WORLD PHOTOS. REPRODUCED BY PERMISSION.

VIRTUAL TENNIS

The ongoing explosion in computer power makes possible the crafting of "virtual" worlds in which a gameplayer, scientist, or other user can experience the illusion of movement and exploration. In most virtual-world or virtual-reality systems, a headset replaces the scene that the user would otherwise see with computer-generated scenes.

But the sense of touch is not so easy to fool. One approach, in virtual tennis, is to have a computer read position information from a racket grip held in the player's hand. A rod is also attached to the "racket." When the player sees a ball coming in the virtual world which the headset shows to them, they swing at the ball. The computer senses the forces that the player's hand exerts on the racket grip, as well as the position of the racket in three-dimensional space, and calculates whether they player is going to succeed in hitting the virtual ball. If they do, the computer sends a shock along a rod connected to the player's racket so that they can feel the impact of the ball—which does not physically exist—hitting the racket. Such systems are already becoming commercially available.

All the computations performed by the computer in such a game involve vectors and linear mathematics. The position of the racket in space is characterized by a set of three-dimensional vectors; the force of the player's grip, the velocity of the ball, and other variables are also represented by vectors. Furthermore, the computer must calculate what racket positions are "feasible" for the system, that is, what positions the rod and wiresattached to the racket can allow. This is done using matrix algebra.

LINEAR PROGRAMMING

Linear inequalities (see sidebar) are important to the problem-solving method known as "linear programming." In a linear programming problem, linear equalities and linear inequalities are combined into a system (that is, they all involve the same variables or unknowns).

Maximizing Profits

The technique known as "linear programming" combines linear equations with linear inequalities (see sidebar) to find the best way of using limited resources. It is used mostly by large organizations, such as corporations or the military, to minimize operating costs.

Banks use linear programming to process checks more efficiently. In particular, they wanted to minimize "float." Float is the amount of money represented by uncancelled checks—checks that have been received by the bank but for which the money has not yet been collected. Float is detrimental to profit because it represents money in limbo; the bank cannot make money *on* that money (invest it) until the check has cleared.

What should a bank do to minimize float without spending so much doing it that the cure is worse than the disease? When checks are received they are "encoded," that is, marked with magnetic ink by a machine. This is the first step in clearing the check. Banks realized they needed to encode checks as quickly as possible without hiring too many machines and clerks, so mathematicians and computer specialists set up a linear programming problem to model the situation. That is, they organized float, encoding machines costs, wages and hours for clerks, and other relevant variables as a set of linear equations and inequalities, and solved this system using linear algebra. The solution showed banks how many full- and part-time clerks to assign to how many shifts on how many machines in order to minimize float. Although there are increasingly high-tech ways to digitize information and handle checks, many financial institutions still use linear programming to save money and increase profits.

This system is then solved, using the methods of linear algebra, to find the "optimum" (i.e., best possible) way of mixing ingredients, manufacturing items, transporting supplies, or allotting other resources.

The first step in a linear programming problem is to define a linear equation that describes something which we want to minimize (expenses, say) and as many linear inequalities as we need to describe the bounds on our resources: for instance, that there are only so many minutes in a day, or pounds of Ingredient Z available, or dollars in the bank available for investment. Each linear inequality is then converted into a linear equality. For example, the inequality $50x_1 + 12x_2 \leq 100$ really says that $50x_1 + 12x_2$ is less than 100 by some unknown amount (maybe 0). This is the same as saying that an $50x_1 + 12x_2$ plus an unknown quantity equals 100. If we name this third unknown M, we can turn the inequality into an equation: $50x_1 + 12x_2 + M = 100$. When all linear inequalities have been turned into linear equations, we then use matrix algebra methods (which are described in many textbooks) to solve the system and find out the best way to run our business.

Linear programming is used by real-life organizations, especially businesses and the military. An example is the use of processing stations in semiconductor manufacturing plants. These plants make the circuit-covered "chips" that run all complex electronic devices, including computers. Many thin layers of material have to be built up on each chip, and each layer requires many stages of optical and chemical processing. In fact, more resources are consumed in making the tiny chips in a desktop computer than in making all the rest of the computer put together. Manufacturers are therefore keen to use their chip-making factories efficiently.

A processing station in a chip factory is a large, complex device that performs one step at a time in the chip-making process. Instead of having hundreds of stations, one for every step, it is cheaper to re-fit each station (change some of its parts) occasionally so that it can do a different step. But refitting a station takes time; it would be unprofitable to refit a station every single time it performed a step. How many batches of chips should a station process before being refitted for another step? Linear programming is used to answer this question, telling the manufacturer how to schedule steps and stations for maximum profit.

LINEAR REPRODUCTION OF MUSIC

If a musician plays two notes in a recording studio, one twice as loud as the other, you want two notes come out of your stereo's speakers so that the one is twice as loud as the other. If graphed on paper, this relationship between live performance and ideal playback is a straight line—a linear function. A great deal of mathematical design work goes into making sound-reproduction systems as linear as possible.

But nonlinearity—electronic behavior that is *not* linear—has its uses, too. The rough sound of a rock guitar is produced by feeding an electrical signal derived from the guitar's strings into a circuit that does not respond linearly. That is, the original signal looks like a complicated wave or series of up-and-down wiggles; when two

Key Terms

Linear algebra: Includes the topics of vector algebra, matrix algebra, and the theory of vector spaces. Linear algebra originated as the study of linear equations, including the solution of simultaneous linear equations. An equation is linear if no variable in it is multiplied by itself or any other variable. Thus, the equation $3x + 2y + z = 0$ is a linear equation in three variables.

Linear equation: An equation whose left-hand side is made up of a sum of terms, each of which consists of a constant multiplying a variable, and whose right-hand side consists of a constant.

Linear programming: A method of optimizing an outcome (e.g., profit) defined by a linear equation but constrained by a number of linear inequalities. The inequalities are recast as linear equation and the resulting system is solved using matrix algebra.

Matrix: A rectangular array of variables or numbers, often shown with square brackets enclosing the array. Here "rectangular" means composed of columns of equal length, not two-dimensional. A matrix equation can represent a system of linear equations.

System of equations: A group of equations that all involve the same variables.

Vector: A quantity consisting of magnitude and direction, usually represented by an arrow whose length represents the magnitude and whose orientation in space represents the direction.

wiggles, one twice as big as the other, are fed into a nonlinear circuit, the larger wiggle does not come out twice as big but gets flattened or chopped off at the top and bottom. This happens because the circuit cannot produce a signal above or below a certain limit. The resulting sound is, technically speaking, "distorted"—but sometimes, that's exactly what we want.

Where to Learn More

Books

Budnick, Frank S. *Finite Mathematics with Applications.* New York: McGraw-Hill, 1985.

Lay, David C. *Linear Algebra and its Applications,* 2nd ed. New York: Addison-Wesley, 1999.

Logarithms

Overview

A logarithm is the power to which a number (usually termed the base number) must be raised to equal a target number.

Fundamental Mathematical Concepts and Terms

In base 10 systems, 2 is the logarithm of 100 because $10^2 = 10 \times 10 = 100$. The number 2 in this example is the exponent of the base number 2 that yields 100. Accordingly, in base 10 the log of 100 is 2.

Because logarithms are so common in mathematics, there are different ways to develop an understanding of them and this will often cause some confusion. However, at their most basic level they can be thought of as a set of rules that allow one quantity to be converted into another to simplify a problem. This idea is the same as multiplication, which is a simplification of the operation of repeated addition; it is easier to say 50×5 rather than write $50 + 50 + 50 + 50 + 50$. Logarithms are effectively the next step, the simplification of repeated multiplication or division. Logarithms have their own form of mathematical notation that can only be manipulated in strict accordance with a set of rules. Once these rules are understood the work of manipulating logarithms is carried by the notation itself, operations no more complex than multiplication, addition, subtraction and division are used to manipulate terms in an equation and generate the desired result.

Before trying to use the mathematical notation of logarithms, it is helpful to understand some aspects of mathematical notation itself. Consider the simple operation of multiplication. The common notation for multiplication is the $\times$ symbol.

Using a symbol to stand for repeated addition, called multiplication, is a form of shorthand. For example, $2 \times 3 = 2 + 2 + 2$. Obviously, the notation reduced the amount of work needed to express repeated addition; it would be hard work if you had to write out $3{,}200 \times 563$ as $3{,}200 + 3{,}200 + 3{,}200 \ldots$ some five hundred sixty three times. However, this notation is more than just a shorthand; it allows us to manipulate quantities that were not possible before and extend our range of mathematical tools. For example, consider multiplying two fractions together. Even though it is not possible to write this out directly, it is still possible to find the answer, $0.5 \times 0.7 = .35$. It is even possible to throw away the numbers and replace them with letters that represent any

number that you can think of. Here x means any number we can think of multiplied by three. y now represents the answer, $3 \times x = y$. In equations like this the multiplication symbol, $\times$, is often dropped and letters and numbers that are next to each other are understood to be multiplied. If we set $x = 2$, remembering this is just a number used for example, our equation y is found to be equal to $3 \times 2 = 6$. Again, if $x = 5$, then y is equal to 15.

One of the great powers of this notation is that it allows the terms, such as x and y, to change places. This is done by noticing that any operation performed to on one side of the equal sign must be repeated on the other side of the equal sign. This is because the values are equal and what we do to one side should balance the effect on the other side.

Let's consider an example of this for the formula $3 \times x = y$. This formula may give us some property of a material, and we conduct an experiment where we measured the values of y of that material. Can we find the value of x? Yes, simply find the value of x by dividing both sides by three: $(3 \times x)/3 = y/3$ so this gives $x = y/3$. By dividing the equation by 3 on both sides we have eliminated the term in front of the x and given the equation in terms of y. This process is called rearranging an equation.

THE POWER OF MATHEMATICAL NOTATION

As multiplication was the extension of repeated addition, so raising to powers, or simply powers, is the extension of repeated multiplication. Consider the number 5 multiplied by itself four times; this can be written in shorthand by putting the number of times the multiplication is to be repeated as a smaller number 4 to the top right of the 5. Here are some examples, $5 \times 5 \times 5 \times 5 = 5^4$ is the same as $5 \times 5 \times 5 \times 5 = 5^4$. To read this notation out loud we say base five raised to the power of four. Another example is $2 \times 2 \times 2 \times 2 \times 2 \times 2 = 2^6$ read as base two raised to the power of six. There are a couple of points to remember about this notation. The first is that the power is also known as the exponent and raising to a power is also known as exponentiation. Exponentiation is not to be confused with the exponential function, e^x, discussed later. Another point to note is that numbers raised to the power of two or three are often read as squared or cubed. For example, $5 \times 5 = 5^2$ is read as five squared and $8 \times 8 \times 8 = 8^3$ is read as eight cubed.

As division is the opposite of multiplication, logarithms can be thought of as the opposite of exponents. They come in two common forms. The first form is written as $\log_{10}$, read as log base 10. The base here is related to the base of the powers, as we shall soon see. Log_{10} is so common that in texts and the buttons on most calculators the base 10 is dropped and it simply reads as log or lg. The other form is read as the natural logarithm, and is written as ln, this is identical to "log e." Logs to any other base, such a base 2, are written as $\log_2$, etc.

POWERS AND LOGS OF BASE 10

Now students should try to get a feel for some values in base 10. Using a scientific calculator, they can try the following, log (1,000) = 3. This tells us that 1,000 can be repeatedly divided by 10 three times: 1,000 /10 /10 /10 = 1, which is true. Another way to look at this is the logarithm has told us that the number 1000 has three zeros after the one. Now raise base 10 to the power of 3 and we are back with the number we started from, $10 \times 10 \times 10 = 10^3 = 1{,}000$. Here we see the relation between the base of the logarithm and the base of the power. This reflects the relationship of logarithms as repeated division and powers as repeated multiplication. Raising the logarithm to the power like this is called an anti-logarithm, and it gives us back the number we started with. Here is another example, log (10,000) = 4. Again, this shows us that 10,000 can be repeatedly divided by 10 exactly four times, or to view it another way, there are four zeros after the 1. Raising this logarithm to the power of base 10 gives us back our number, $10^4 = 10{,}000$.

For any number made from one followed by a number of zeros the log will always equal the number of zeros if we use logarithms with base 10.

As with the previous multiplication example, our definitions of this notation allow us to extend this idea of repeated multiplication and division to more than just shorthand, because we can now use fractional values. Students should try the following, log (5,246) = 3.7198283. Even though this cannot be written out as an exact repeated division by ten as we did before, it still tells us how 5,246 would divide into 10 in an abstract sense, about four times. If we raise this to base ten do we get the answer back as expected? $10^{3.7198283} = 5{,}246.0002$. Almost, but what about the small fraction after the number? (Depending on places and the calculator, the exact fraction may differ.) The digits after the decimal place are not important and are there because the calculator cannot store numbers to infinite precision. However, they can safely be ignored as the error is not in the digits in which we are interested. This will always be found to be true and we obtain the correct answer of 5,246. The notation has allowed the extension of the mathematical idea of repeated multiplication to be taken beyond the simple idea of a shorthand.

LOGARITHMS TO OTHER BASES THAN 10

What about logarithms with a base other than 10, such as, $\log_2 (256) = 8$? You do not find a $\log_2$ button on your calculator because logarithms to bases other than 10 can always be expressed as $\log_{10}$ using the following formula:

$$\log N (y) = \log_{10} (y)/\log_{10} (N)$$

Here y is the value of the log and N is the value of the base. So, to solve the previous equation, $\log_2 (256) = \log_{10} (256)/\log_{10} (2) = 2.40824/0.30103 = 8$. As a check, $2^8 = 256$, as expected.

Logarithms to the base 2 are common in computing where a computer will represent numbers by a series of 1s or 0s internally. Arithmetic performed in base two is called binary.

POWERS AND THEIR RELATION TO LOGARITHMS

Let us consider replacing the numbers with letters as was done with multiplication. Again the letter x can take any value, $y = 10^x$. If we apply log to the terms on both sides of the equals sign we can now find x. Check the method used for rearranging the formula, $3x = y$, if you do not understand this step, $\log (y) = \log (10^x) = x\ x = \log (y)$. This shows the effect of the log was to cancel the base of the power, 10, in $y = 10^x$. This is the same in any base and generally can be written $\log_N (N^x) = x$. Notice that the base must be the same in both parts. For instance, the following formula is wrong, $\log_2 (3^x)$ is not equal to x but this is correct, $\log_3 (3^x) = x$. This rule allows us to cancel the base of a power by multiplication with a logarithm. This is useful for extracting the power x.

THE ALGEBRA OF POWERS AND LOGARITHMS

When two powers of the same base are multiplied, the repeated multiplication is effectively extended. This is identical to the base raised to the sum of the powers. If they are divided, then the powers are just subtracted. $N^x \times N^y = N^{(x+y)}$ $(N^x)/(N^y) = N^{(x-y)}$.

While thinking about these relations, students can see that we have combined multiplication and addition and vise-versa for division and subtraction. Using the logarithm to extract the powers, shown previously, allows the addition and subtraction parts to be extracted, $\text{Log}_N (N^x \times N^y) = x + y = \log_N (N^x) + \log_N (N^y)$ and if we set $N^x = A$ and $N^y = B$ then $\log_N (A \times B) = \log_N (A) + \log_N (B)$. Likewise $\log_N (A/B) = \log_N (A) - \log_N (B)$. The rules shown here are the reason that logarithms allow us to reduce the complexity of large lists of multiplications (or divisions) down to simple addition (or subtraction).

These basic properties of logarithms were critical in the development of science and industry over the past three centuries.

LOG TABLES

Suppose you want to multiply numbers so large that it will take a while to complete the computation by hand. It would be faster if there were some sort of table to look up the answer. If we wanted the answer to be accurate to four digits, a simple solution would be to make a table, called a matrix, with the rows and columns corresponding to the numbers between 1 and 9,999. If we picked a row, say 50 and column say 26, were they crossed we would find the answer for the multiplication, in this case 1300. Picking a row and column would show us the multiplied answer quickly.

Now any multiplication can simply be looked up in our table. For numbers outside the range 1 to 9,999 we can still find the answer by moving the decimal point until they are in this range, reading from the table and finally moving the decimal place back by an opposite number of steps.

The problem with this basic system, that makes it unworkable, is the number of entries needed will be huge. For four digits accuracy each edge of our square table would have 10,000 numbers. This gives us least 100,000,000 entries ($10{,}000^2$). If the print is very small that is still enough to fill 20 thick books.

Another problem is seen as we increase the number of digits accuracy needed. The square shape of our table, the matrix, will rapidly start to get bigger with each digit added. Most scientific and engineering calculations work at seven digits accuracy. This works out to be more than a million thick books to store our table and we have not even considered division.

History of Logarithms

Logarithms were invented in the seventeenth century by John Napier, a Scottish Barron. During that time in Scottish history the country was undergoing major religious and political upheaval. In this climate academic study was not held in high regard. Later in life Napier considered his greatest publications were his theological works, with his mathematical works as a secondary interest. The development of logarithms at this time came from a need to simplify the computations of repeated multiplications and divisions. These computations were

"e"

The most common logarithms that are encountered in mathematics are natural logarithms. These are logarithms to the base e. This is due to the remarkable properties of the number e and its many special properties. One of these properties is that e^x has a rate of change that is equal to its value of x. This makes e^x a solution to equations used to calculate rates of change, often with respect to time, called differential and integral equations. This number is irrational, which means that the numbers after the decimal place carry on forever and the sequence never repeats. The first five significant digits of this number are 2.7182. e has a very special place in mathematics and is believed to be a fundamental number in nature.

Applications of e are too numerous to list, but some examples are the calculation of compound interest rates, rates of radioactive decay, or the rates of decay of damped springs found in the suspension systems of cars. This is why a scientific calculator will have the natural logarithm button ln, as this is the most common logarithm encountered in engineering, scientific, or mathematic work.

One example of the families of equations that contain this number are said to show exponential behavior. This means that they can rapidly change with time. This behavior is seen in many natural and human-made systems. Some examples are the rates of growth of bacteria and radioactive decay, and the calculation of compound interest rates.

common in the calculations of astronomical charts used by the navy and the shipping industry and religious charts used by the church, three of the most powerful institutions in Britain at the time.

The power of the system of logarithms comes from its simplification of computational steps involved. By converting terms that were to be multiplied or divided to logarithms from a table, they could then simply be added or subtracted, and the result read from another table. The system was compact and flexible enough that the two tables needed to perform the steps could be listed in a few pages at the back of a book. Another device, called a slide rule, reduced computational time further by allowing the user to read the answer directly by simply moving the two rulers on the device. For 300 years this device was commonly used by scientists and engineers, just as hand-held calculators are today.

It is no understatement that the invention of logarithms changed the world. Their usefulness in industry and science was soon realized, and the system rapidly spread around the world. The invention of the electronic calculator in the 1950s allowed complex calculations to be performed by the simple push of a button.

Real-life Applications

COMPUTER INTENSIVE APPLICATIONS

Although the system of using log tables is not in common use since the invention of electronic calculators it has found new life in computer intensive applications. The desktop calculator will have to run a program to multiply numbers together. This process is so fast we cannot see it and to us it looks instant. However, it still takes a certain amount of time, and the time needed increases with the complexity and amount of calculations. In a computer-intensive application, in which millions of numbers have to be multiplied every second and speed is critical, this can become a problem. Some examples are interactive 3D computer games, and software used in spacecraft and aircraft. Here, the small time the computer takes to calculate the numbers will rapidly increase and can become substantial.

This can be reduced if a set of log and anti-log tables are wired into the computers memory, and the computer only has to look up values instead of running a program to find the answer. This technique to improve performance under heavy arithmetic load is called a log lookup table.

USING A LOGARITHMIC SCALE TO MEASURE SOUND INTENSITY

Decibels (dB) are used as a measure of sound level. They are common markings for stereos, televisions, and other audio equipment and are based on a log10 scale. The faintest sound we can hear is called the threshold of hearing. Its value is tiny, about a 0.3 billionths change in air pressure. The scale is given as dB = log (Number of times greater than threshold of hearing) $\times$ 10. A normal conversation is 60dB or, remembering to divide by the 10 from right of the formula, 10^6 = 1,000,000. This is one million times louder than the faintest sound you can hear. We can safely hear sounds to around 90 dB, the level of an orchestra, before damage to the ear starts to occur, but we can still hear sounds louder than that. The levels of the front row of a rock concert can reach 110 dB. After this, there is pain and instant deafness. This gives the human ear an amazing range of about 100 billion times the faintest sound it can detect.

Tourists watch as tsunami waves hit the shore near in Penang, northwestern Malaysia on December 26, 2004. Although the earthquake causing the tsunami initially measured at 8.0 on the Richter scale, scientists later found that the data indicated an undersea earthquake measuring 9.0. The difference of one point is not insignificant. On the logarithmic Richter scale, each whole number increase means that the magnitude of the quake is ten times greater than the previous whole number. Thus, an earthquake with a magnitude of 9.0 has ten times the force of one with a magnitude of 8.0; an earthquake of 9.0 has 100 times the intensity of the 7.0 earthquake, etc. AP/WIDE WORLD PHOTOS. REPRODUCED BY PERMISSION.

ESTIMATING THE AGE OF ORGANIC MATTER USING CARBON DATING

The atmosphere is continuously being bombarded by radiation from space. In the upper atmosphere, the radiation from space has enough energy to change atoms of nitrogen into carbon. Carbon created this way is called carbon 14 and is different to the majority of the carbon we see around us called carbon 12. Carbon 14, unlike carbon 12, is unstable and will slowly decay back to nitrogen over a period of many thousands of years. The rate of production of carbon 14 in the atmosphere can be shown to be stable for a very long period of history, and this allows us to measure the age of dead organic matter.

All life on Earth is made from carbon, and during the course of an organism's life it will absorb small amounts of carbon 14. When the organism dies, it will stop absorbing carbon 14. So by measuring the ratio of carbon 14 to carbon 12 that is present and using the law of exponential decay of a radioactive source and their logarithms, scientists can calculate the age of the material.

DEVELOPING OPTICAL EQUIPMENT

No matter how pure a material is made, as light passes through it a small majority will always be scattered or absorbed. This is an exponential effect and logarithms are therefore used extensively in the design of optical equipment. Just a few examples are cameras, optical fibers, and the design of television screens.

USE IN MEDICAL EQUIPMENT

Certain cancers can be treated by passing radioactive beams though the body. A machine with a radioactive

source is rotated around the patient. Only at the center of this rotation will the radiation be constant. Moving away from the center, the radiation will only pass through the patient periodically as the machine makes each rotation.

The location of the cancer is carefully mapped, and the absorption of the radiation through the body and the absorption by various organs is then calculated. This requires the use of logarithms due to the exponential nature of this absorption. The aim of the surgeons is to locate the cancer and manipulate the intensity of the beam over each rotation so that minimum damage is caused to the surrounding organs and maximum damage is caused to the tumor.

DESIGNING RADIOACTIVE SHIELDING FOR EQUIPMENT IN SPACE

Outside the protection of the Earth's atmosphere we enter a highly radioactive environment. Spaceships, satellites, and spacesuits must all able to absorb and disperse this energy to protect the delicate equipment and astronauts from damage. The absorption must be balanced against weight not too massive to launch.

Absorption of different types of radiation and in different materials involves calculations using logarithms due to the exponential nature of absorption.

SUPERSONIC AND HYPERSONIC FLIGHT

During supersonic and hypersonic flight, the air flow over the craft behaves very differently than at slower flight speeds. Logarithms are used in the design and fuel requirements.

Potential Applications

CRYPTOGRAPHY AND GROUP THEORY

Cryptography is the science of encoding information in such a way that an eavesdropper cannot intercept and decode a message. Modern methods rely on a mathematical phenomenon that some formulas are practically impossible to invert. This means that information encoded by such a formula cannot simply be decoded by rearranging the terms of the formula to reverse the processes.

Two parties can generate and swap unique keys which will unlock the message encrypted by a formula like this. However, if an eavesdropper were to try to decode the encryption by setting a machine up between the two parties without the keys, he would have to invert the formula used to encrypt the information.

One set of such functions that show these properties come from an abstract area of mathematical research that studies the relations between objects called group theory. Certain groups can be given properties that act like exponentials and logarithms. The calculation of the exponential part of these groups is very simple, and the calculation of the logarithmic part is very hard. This property can be exploited in cryptography. Studies of this branch of mathematics are important in the future development of faster and more secure algorithms.

Where to Learn More

Books

Durbin, John R. *College Algebra*. New York: John Wiley & Sons, 1985.

Morrison, Philip and Phylis Morrison. *Powers of Ten: A Book About the Relative Size of Things in the Universe and the Effect of Adding Another Zero*. San Francisco: Scientific American Library, 1982.

Periodicals

Curtis, Lorenzo. "Concept of the exponential law prior to 1900." *American Journal of Physics* 46(9), Sep. 1978, pp. 896–906 (also available at <http://www.physics.utoledo.edu/~ljc/explaw.pdf>.

Web sites

SOS Math! "Introduction to logarithms." <http://www.sosmath.com/algebra/logs/log1/log1.html> (February 1, 2005).

Logic

Overview

Logic is a set of rules by which decisions and conclusions are either derived or inferred from a set of statements. Logic can be mathematical or predicate (dealing with statements and sentences). Logic is also a set of rules by which computers handle data, and circuit logic dictates how many devices operate.

However, a logical decision or a belief may or may not be correct. Logic is more of a set of rules to follow in reaching a decision. An example of a logical, but incorrect, bit of reasoning is the following: If I believe that sheep have a wool coat and that all sheep are mammals, then it could make logical sense for me to believe that all mammals have wool coat. The conclusion is incorrect, but it is logically drawn.

Fundamental Mathematical Concepts and Terms

Over twenty-four centuries ago, the idea of logic was explored and developed about the same time in China, India, and Greece. The Greek philosopher Aristotle (384 B.C.–322 B.C.) was important in the creation of logical systems.

REASONING

Logic does not necessarily lead to the truth. What logic does do is to allow us to look at an argument and to decide if the reasoning is valid or not valid. Logic also points out how we can come to believe something that is not true (even though that sounds illogical).

PROPOSITION AND CONCLUSION

The starting point of a logical line of thought is called the proposition (or the statement). A proposition is the real meaning of the sentence (or the equation, as it can be written in mathematical language also). The meaning can be expressed in different ways and still mean the same thing. For example "Today is Friday" and "Yesterday was Thursday" are the same proposition, while "My name is Brian" is a different proposition.

A proposition is always true or false, although it is sometimes unknown which proposition is true and which is false. "There is life on Mars" is an example of a proposition that may or may not be true; we have yet to find out.

Logic proceeds from the starting point of the proposition to the conclusion in a series of steps that are related

Proposition	Steps	Conclusion
True	Support the conclusion	Always true
True	Do not support the conclusion	Can be true or false
False	Support the conclusion	Can be true or false
False	Do not support the conclusion	Can be true or false

Table 1.

to each other. That is, one step is followed by a step that supports it.

Here is an example of a logical series of steps:

- Today is Friday.
- My library books were due Thursday.
- My library books are overdue.

Here is an example of a series of steps that is not logical:

- The moon is full.
- There are clouds in the sky.
- My cat has a hairball.

From the proposition, the steps that proceed to the conclusion can be set up so that the steps guarantee that the conclusion is true. This is a good style to use when debating. There is no middle ground with this type of approach. Either all the steps lead to a single conclusion or they do not.

A number of different outcomes can still result, depending on whether the steps from the proposition to the conclusion support this conclusion. Table 1 summarizes these various possibilities.

A less rigid style is when the steps from the proposition to the conclusion support the likelihood of the conclusion. In this style a conclusion does not have to be true, it is just likely to be true. Points can be presented that support the conclusion, but the conclusion could still be debatable. This style of logic is used in many courtrooms by lawyers trying to defend their clients from charges brought against them.

Real-life Applications

BOOLEAN LOGIC

Many persons do most of their banking while sitting at their desk. This is possible since they can hook up to the local bank's Web site, research bank accounts, and then use the computer directions built into the site to shift money from one account to another, pay bills, and look at the action in each account over whatever time period is desired.

These activities are pretty human-like. How can computers do them? The answer is something called Boolean logic.

Boolean logic is named after the Irish mathematician George Boole (1815–1864). From an early age, Boole showed a talent for languages and teaching. When he was 20, Boole began to teach himself mathematics. He proved to be talented at this as well, publishing papers in the leading math journals of the day. When he was 34 years old, he was appointed chair of mathematics at Queens College in Cork, Ireland. He taught there for the rest of his life.

In 1854, when he was only 39 years old, Boole published a paper called "An Investigation into the Laws of Thought, on Which are founded the Mathematical Theories of Logic and Probabilites." The ideas in this paper became the basis of Boolean logic.

One niche that Boolean logic has filled beautifully is the task of sifting through vast amounts of information to find those bits of information that are desired.

FUZZY LOGIC

Fuzzy logic is a way of making computers behave in a way that is similar to the way humans think. Often, we are able to use information that is not really clear or precise to make decisions that are definite. We can relate the imprecise (fuzzy) information with what we already know to make a decision.

Here is an example. You are driving your car on a crowded, four-lane freeway. The speed limit is 65 mph (105 km/h). As is usually the case, traffic is moving faster, at an average speed 70 mph (113 km/h). You know that it

Boolean Logic and Computer Searches

Boolean logic links the common parts of different pieces of information. This feature makes Boolean logic widely used in Internet search engines. For example, if there was no Boolean logic and information from the Internet on the trigonometry and homework problems was desired, the Internet search for every word would show all the documents that separately mention "trigonometry" or "homework." This would probably result in a huge number of sites to search, making the search nearly meaningless. Because of Boolean logic, however, a search can be done to look for those documents that contain "trigonometry" AND "homework." This number of sites will be much less, and the sites will be more likely to have something to do with homework related to trigonometry rather than homework related to all subjects.

Boolean logic even allows a search to focus on one word and not another. To use the above example, the following search could be done: "trigonometry" AND "homework" NOT "advanced." This would allow the search engine to zero in on those site that were about teaching methods of trigonometry at a basic level as opposed to sites that discussed advanced trigonometry.

is safest for you and those around you to drive "with the traffic." But what exactly does driving "with the traffic" mean?

Watching other drivers, you realize that driving "with traffic" is done different ways. Some drivers will drive more slowly and stay in the right hand lane. Other drivers will speed and zig-zag their way between cars and lanes. Usually, the different styles mesh together to make a smooth flow of traffic. When they do not, there a traffic accident can occur.

Fuzzy logic was conceived by Lotfi Zadeh, a professor of electrical engineering at the University of California at Berkley, and was first proposed in a 1965 paper. From its humble beginnings, fuzzy logic has expanded to assume an important role in our daily lives. For example, because of fuzzy logic, the computer inside a video camera is able to keep focusing even when the camera is jostled. As another example, fuzzy logic makes it possible to program a microwave oven to cook differently sized and types of foods perfectly with the touch of one button.

The logic of fuzzy logic can be summed up as IF X AND Y THEN Z. It is the 'if' and 'and' that makes things less precise.

The following example may help to make this fuzziness clearer. A conventional oven operates on the basis of exact temperature. A thermometer in the oven can cut off the power to the oven's heater when the oven reaches whatever temperature has been selected, and will kick the heater back into action when the temperature falls below another set value. This occurs no matter what is in the oven.

A microwave with a fuzzy logic temperature control does not rely on exact temperatures. Instead, the process is like this: "IF (the process is too cool) AND (the process is getting colder) THEN (add more heat)", or "IF (the process is too hot) AND (the process is getting colder) THEN (heat it up now)."

Companies have leapt on fuzzy logic as a way of making products that will perform better for people. Self-focusing cameras and video recorders, washing machines that can adjust the strength of cleaning power to how much dirt is in the clothes being washed, the controls to car engines, anti-lock braking systems in vehicles, banking programs, programs that allow people to do stock market trades—all these would not exist if not for fuzzy logic.

Where to Learn More

Books

Bennett, D.J. *Logic made Easy: How to Know When Language Deceives You.* New York: W.W. Norton & Company, 2004.

Gregg, J.R. *Ones and Zeros: Understanding Boolean Algebra, Digital Circuits, and The Logic of Sets.* New York: Wiley-IEEE Press, 1998.

Mukaidonon, M., and H. Kikuchi. *Fuzzy Logic for Beginners.* Singapore: World Scientific Publishing Company, 2001.

Web sites

Brain, M. "How Boolean Logic Works." <http://computer.howstuffworks.com/boolean.htm> (September 2, 2004).

Cohen, L. "Boolean Searching on the Internet: A Primer in Boolean Logic." *University Libraries-State University of New York.* <http://library.albany.edu/internet/boolean.html> (September 3, 2004).

Kemerling, G. "Arguments and Inference" <http://www.philosophypages.com/lg/e01.htm> (September 3, 2004).

Matrices and Determinants

Overview

In mathematics, a matrix is a group of numbers that have been arranged in a rectangle. The word for more than one matrix is matrices. The mathematics of handling matrices is called matrix algebra or linear algebra. Matrices are one of the most widely applied of all mathematical tools. They are used to solve problems in the design of machines, the layout by oil and trucking companies of efficient shipping routes, the playing of competitive "games" in war and business, mapmaking, earthquake prediction, imaging the inside of the body, prediction of both short-term weather and global climate change, and thousands of other purposes.

Fundamental Mathematical Concepts and Terms

Matrices are usually printed with square brackets around them. The matrix depicted in Figure 1 contains four numbers or "elements."

A column in a matrix is a vertical stack of numbers: in this matrix, 3 and 7 form the first column and 5 and 4 form the second. A row in a matrix is a horizontal line of numbers: in this matrix, 3 and 5 form the first row and 7 and 4 form the second. Matrices are named by how tall and wide they are. In this example the matrix is two elements tall and two elements wide, so it is a 2×2 matrix. The matrix in Figure 2 is 3 elements tall and five elements wide, so it is a 3×5 matrix.

A flat matrix that could be written on the squares of a chessboard, like these two examples, is called "two dimensional" because we need two numbers to say where each element of the matrix is. For instance, the number "10" in the 3×5 matrix would be indicated "row 2, column 4." A matrix can also be three-dimensional: in this case, numbers are arranged as if on the squares of a stack of chessboards, and to point to a particular number you have to name its row, its column, and which board in the stack it is on. There is no limit to the number of dimensions that a matrix can have. We cannot form mental pictures of matrices with four, five, or more dimensions, but they are just as mathematically real.

The numbers in a matrix can stand for anything. They might stand for the brightnesses of the dots in an image, or for the percentages of spotted owls in various age groups that survive to older ages. In one of the most important practical uses of matrices, the numbers in the matrix stand for the coefficients of linear equations. A linear equation is an equation that consists of a sum of

variables (unknown numbers), each multiplied by a coefficient (a known number). Here are two linear equations: $2x + 3y = 11$ and $7x + 9y = 0$. Where the variables are x and y and the coefficients are the numbers that multiply them (namely 2, 3, 7, and 9). Together, these two equations form a "system." This system of equations can also be written as a 2×2 matrix times a 2×1 matrix (or "vector"), set equal to a second 2×1 matrix, as depicted in Figure 3.

The information that is in the matrix equation is also in the original system of equations, and is in almost the same arrangement on the paper. The only thing that has changed is the way the information is written down. For very large systems of equations (with tens or hundreds of variables, not just x and y), matrix equations are much more efficient.

Say that we wish to solve this matrix equation. This means we want to find a value of x and a value of y for which the equation is true. In this case, the only solution is

$$x = -33,\ y = \frac{77}{3}$$

(If you try these values for x and y in the equations $2x + 3y = 11$ and $7x + 9y = 0$, you'll find that both equations work out as true. No other values of x and y will work.) Finding solutions to matrix equations is one of the most important uses of computers in science, engineering, and business today, because thousands of practical problems can be described using systems of linear equations (sometimes very large systems, with matrices of many dimensions containing thousands or millions of numbers). Computers are solving larger matrix equations faster and faster, making many new products and scientific discoveries possible.

The rules for doing math with matrices, including solving matrix equations, are described by the field of mathematics called "matrix algebra." Matrices of the same size can be added, subtracted, or multiplied. One number that can be calculated from any square matrix—that is, any matrix that has the same number of rows as it has columns—is the determinant. Every square matrix has a determinant. The determinant is calculated by multiplying the elements of the matrix by each other and then adding the products according to a certain rule. For example, the rule for the determinant of a 2×2 matrix is as follows:

$$\begin{bmatrix} a_1 & b_1 \\ a_2 & b_2 \end{bmatrix} = a_1b_2 - a_2b_1$$

$$\begin{bmatrix} 3 & 5 \\ 7 & 4 \end{bmatrix}$$

Figure 1: A matrix with four numbers or "elements."

$$\begin{bmatrix} 5 & 9 & 4 & 2 & 5 \\ 3 & 1 & 0 & 10 & 2 \\ 9 & 9 & 5 & 2 & 7 \end{bmatrix}$$

Figure 2: A 3×5 matrix.

$$\begin{bmatrix} 2 & 3 \\ 7 & 9 \end{bmatrix} \times \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 11 \\ 0 \end{bmatrix}$$

Figure 3.

(Here the small numbers attached to the variables are just labels to help us tell them apart). For a 3×3 matrix, the rule is more complicated:

$$\begin{bmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{bmatrix} = a_1b_2c_3 + a_3b_1c_2 + a_2b_3c_1 - a_1b_3c_2 - a_2b_1c_3 - a_3b_2c_1$$

—and the rules get more and more complicated for larger matrices and higher dimensions. But that isn't a problem, because computers are good at calculating determinants.

Determinants are always studied by students learning matrix algebra, where they have many technical uses in matrix algebra. However, they are less important today in matrix theory than they were before the invention of computers. About a hundred years ago, a major mathematical reference work was published that merely summarized the properties of determinants that had been discovered up to that time: it filled four entire volumes. Today, mathematicians are less concerned with determinants than they once were. As one widely used textbook says, "After all, a single number can tell only so much about a matrix."

Real-life Applications

DIGITAL IMAGES

A digital camera produces a matrix of numbers when it takes a picture. The lens of the camera focuses an image

Key Terms

Matrix: A rectangular array of variables or numbers, often shown with square brackets enclosing the array. Here "rectangular" means composed of columns of equal length, not two-dimensional. A matrix equation can represent a system of linear equations.

Vector: A quantity consisting of magnitude and direction, usually represented by an arrow whose length represents the magnitude and whose orientation in space represents the direction.

on a flat rectangular surface covered with tiny light-sensitive electronic devices. The devices detect the color and brightness of the image in focus, and this information is saved as a matrix of numbers in the camera's memory. When a picture is downloaded from a camera to a computer and altered using image-editing software, it is subjected to mathematical manipulations described by matrix algebra. The picture may also be "compressed" so as to take up less computer memory or transmit over the Internet more quickly. When an image is compressed, the similarities between some of the numbers in its original matrix are used to generate a smaller matrix that takes up less memory but describes the same image with as little of image sharpness lost as possible.

FLYING THE SPACE SHUTTLE

In the early days of flight, pilots pushed and pulled on a joystick connected to wires. The wires ran over pulleys to the wings and rudder, which steered the plane. It would not be possible to fly as complex a craft as the Space Shuttle, which is steered not only by movable pieces of wing, but by 44 thruster jets, by directly mechanical means like these. Steering must be done by computer, in response to measurements of astronaut hand pressure on controls. In this method the flight computer combines measurements from sensors that detect how the ship is moving with measurements from the controls. These measurements are fed through the flight computers of the Shuttle as vectors, that is, as $n \times 1$ matrices, where the measurements from ship and pilot are the numbers in the vectors. The ship's computer performs calculations on these vectors using matrix algebra in order to decide how to move the control surfaces (moveable parts of the wing and tail) and how to fire the 44 steering jets.

POPULATION BIOLOGY

One of the things that biologists try to do is predict how populations of animals change in the wild. This is known as the study of population dynamics because in science or math, anything that is changing or moving is in a "dynamic" state. In population biology, a matrix equation describes how many members of a population shift from one stage of their reproductive life to the next, year to year. Such a matrix equation has appeared in the debate over whether the spotted owl of the Pacific Northwest (United States) is endangered or not. If the numbers of juvenile, subadult, and adult owls in year k are written as J_k, S_k, and A_k, respectively (where the small letters are labels to mark the year), and if the populations for the next year, year $k + 1$, are written as J_{k+1}, S_{k+1}, and A_{k+1}, then biologists have found that the following matrix equation relates one years' population to the next:

$$\begin{bmatrix} 0 & 0 & .33 \\ .18 & 0 & 0 \\ 0 & .71 & .94 \end{bmatrix} \begin{bmatrix} J_k \\ S_k \\ A_k \end{bmatrix} = \begin{bmatrix} J_{k+1} \\ S_{k+1} \\ A_{k+1} \end{bmatrix}$$

By analyzing this equation using advanced tools of matrix algebra such as eigenvalues, biologists have shown that if recent rates of decline of habitat loss (caused by clearcutting) continue, the spotted owl may be doomed to extinction. Owls, like all predators, need large areas of land in which to hunt—for spotted owls, about 4 square miles per breeding pair.

DESIGNING CARS

Before the 1970s, car makers designed new cars by making first drawings, then physical models, then the actual cars. Since the 1970s, they have also used a tool called computer-aided design (CAD). CAD is now taught in many high schools using software far more sophisticated than was available to the big auto makers in the beginning, but the principles are the same. In automotive CAD, the first step is still a drawing by an artist using their imagination—a design for how the car will look, often scrawled on paper. When a new image has been agreed on, the next step is the creation of a "wireframe" model. The wireframe model is a mass of lines, defined by numbers stored in matrices, that

outline the shape of every major part of the car. The numbers specify the three-dimensional coordinates of enough points on the surface of the car to define its shape. The wireframe model may be created directly or by using lasers to scan a clay model in three dimensions. The wireframe car model is stored as a collection of many matrices, each describing one part. This model can be displayed, rotated, and adjusted for good looks. More importantly, by using matrix-based mathematical techniques called finite element methods, the car company can use the wireframe model to predict how the design will behave in crashes and how smoothly air will flow over it when it is in motion (which affects how much gas the car uses). These features can be experimentally improved by changing numbers in matrices rather than by building expensive test models.

Where to Learn More

Books

Lay, David C. *Linear Algebra and its Applications*, 2nd ed. New York: Addison-Wesley, 1999.

Strang, Gilbert. *Linear Algebra and its Applications*, 3rd ed. New York: Harcourt Brace Jovanovich College Publishers, 1988.

Measurement

Overview

Measurement is the quantifying of an exact physical value. The thing being measured is normally called a variable, because it can take different values in different circumstances.

The most common type of measurement is that of distance and time. As science and mathematics have developed, accuracy has increased dramatically in both microscopic and macroscopic scenarios. Without the ability to measure, many of the things that societies take for granted simply would not exist. Scientists can now accurately measure quantities ranging from the height of the tallest mountains, to the force of gravity on a distant moon through to the distance between stars light years away. The list is practically endless. Yet, the most remarkable aspect is that much of this can be achieved without a direct physical measurement. It is this theoretical aspect, along with the obvious practical consequences, that makes the area of measurement so fascinating.

Fundamental Mathematical Concepts and Terms

Distance is nearly always defined to be the shortest length between two points. If it were not so defined then there would be an infinite number of distances. To realize this, just plot two points and attempt to sketch all the different-length curves that exist between them. It would take literally forever.

Time is a much harder item to quantify. There are many philosophical discussions about the nature of time and its existence. In physics, time is the regular passing of the present to the future. This is perfectly adequate, until one starts to travel extremely fast. Believe it or not, time starts to slow down the faster one travels. This theory leads into a whole branch of science known as relativity.

The majority of science is also concerned with the measurement of forces, namely the interactions that exist between different objects. The most common forces are gravity and magnetism. The ability to measure these has immediate consequences in terms of space travel and electricity, respectively, as well as the interaction and motion of objects.

Finally, the measurements of speed, velocity, and acceleration have profound implications on travel. Velocity is defined to be the speed of an object with the direction of travel specified. Acceleration the rate at which velocity is changing. A large acceleration indicates that speed is increasing quickly.

Measurement can also be made of the geometrical aspects of objects. This has led to the development of angle-measuring techniques. Much of the development in navigation and exploration has come from the consequences of measuring geometry accurately.

A Brief History of Discovery and Development

The human race has been assigning measurements for thousands of years. An early unit was the cubit. This was defined to be the distance from a man's elbow to the end of his outstretched middle finger. However, as trade began to develop, the need for a standardized system became more important. It made little sense to have two workers producing planks of two cubits, when the two people would differently define a cubit. Inconsistencies would cause considerable problems in engineering projects.

It is acknowledged that the Babylonians were the first to standardize weight. This was particularly important in trade. They had specified stones of fixed weight and these were used to weigh and, hence, value precious gems and jewels. This led to the common terminology of stone, though even this had a variance; the stone used by a fisherman being half that used by a wool merchant. Yet, this wouldn't be a problem because the two trades were distinct. Definitions of mass and its measurement were especially important in the trade of gems.

It was King Edward I who first standardized the yard, a measurement used predominantly in Britain. There was an iron bar named the iron ulna from which all yardsticks were derived. This allowed for a standardized measurement. However, the metric system used today was first developed during the 1790s, Napoleon's time, by the French government. The meter was initially defined to be one ten-millionth of the distance from the North Pole to the equator passing through Paris.

It was not until 1832 that any legal standardized lengths existed in the United States. Indeed, it was a bill in 1866 that finally accepted the metric system. Finally, in 1975, Congress passed the Metric Conversion Act and so the metric system became the predominant measuring system in the United States.

Shortest Distance Between Two Cities

When looking for the distance between two cities in geography the temptation is to pick up a two-dimensional map, measure how far apart the cities are with a ruler, and use a suitable scale to evaluate this route. However, this method is completely erroneous. The best way to understand the following idea is to get a globe and actually test the following for your self. Pick two places of your choice. Get a piece of string and lay it on your flat map between the two cities marking where the cities are. Cut the string to this length. Apart from very few cases, you will be able to place the string on the globe so as to find a route that involves less string.

This is a simple example of spherical geometry. An analogous way to think about the difficulties is to imagine you are walking along when confronted by a hill. It is often the case that the shortest route passes around the hill, and not straight over the summit. The globe is a perfect form of a hill. There is a simple was to find the shortest route between two points. The globe can be cut into Great Circles. These are circles that cut the globe exactly in half. The Great Circle that passes through both cities provides the shortest route.

Real-life Applications

MEASURING DISTANCE

It is extremely easy to measure the distance between two objects if they are of close proximity and, perhaps more importantly, if travel can be done easily between these two points. Travel plans are based on such distances, and such factors to consider include buying gas or deciding whether to walk or take public transport.

All companies involved in travel need to know the distance between start and destination. Customers like to know how far they are going so that they can plan for their arrival. On a smaller scale, production companies need to be able to produce items of a specified length. It is impossible to build a building and yet be unable to specify the lengths of materials required. Or, in the production of parts in either an automobile of a plane, it would be rather difficult to manufacture such products if the pieces didn't fit together. It would be as useless as a jigsaw with incorrect pieces.

DIMENSIONS

Travel and objects do not exist in only one dimension. Instead, there are three aspects to most items: length, depth, and height. These are called the three dimensions. It is

Activity: Running Tracks

Most people would state that a running track is defined to be 400 meters. After all, in a 400-meter race they run once around the track. Yet, a confusing fact is that at the start of the race the runners all start spread out. Is it the case that runners on the outside lanes (of which there are normally eight) are given an unfair advantage?

The solution to this comes simply from careful mathematical thought. Indeed mathematics is required by designers to ensure that any running track meets defined standards. A multi-billion dollar Olympic industry depends on such accuracy. Consider the running track to be two straights and two half circles. It is quite clear that on the outer lane, the runner is passing along a greater circle. We can simplify the problem by joining the two half circles together to create one whole circle. If every lane is a fixed distance d apart, then as you progressively move to outer lanes, each runner would have to run an extra $2 \times \pi \times d$. This equates to an approximate distance of six meters per lane; hence the reason for the staggered starts.

therefore easy to specify the size of a brick required in construction by just stating the three measurements required.

Of course it is not always the most efficient way to specify an object. Circular laminas, thin plates, are often used in construction. In this case, the radius is a much better way of defining the shape.

Mathematicians often define different situations by using different systems of dimensions. This is done to make both the mathematics and its application in real life as simple as possible. Dimensions can also be distorted by the real world. For instance, if a person travels from New York to Los Angeles, can they actually travel in a straight line, even ignoring heights above sea level? The simple answer is no, because the actual world itself is curved to form a global shape. This raises a whole new area of mathematics called curvature, which explores how curved surfaces affect distances.

ACCURACY IN MEASUREMENT

Given a specified measurement in a specified dimension, the next task often required is to be able to do such a measurement accurately, often within a distance imperceptible to the naked eye. It is actually impossible to achieve perfect measurement, yet it is possible to get it to within specified bounds. These bound are often referred to as error bounds. The importance of the situation will limit the accuracy required. There is no point in spending millions of dollars to get a kitchen table an exact length to within on thousandth of a millimeter, yet it is essential that such money be spent when the context is building planes or space shuttles. If there were substantial errors, then the consequences could be disastrous.

Clearly such accuracy cannot be achieved using the simple ruler. It is virtually impossible to measure to within one millimeter without using magnification techniques. It was Sir Isaac Newton who, in 1672, stumbled upon a method called interferometry, which is related to the use of light to accurately measure at microscopic level.

EVALUATING ERRORS IN MEASUREMENT AND QUALITY CONTROL

Being aware of the error involved in real-life production can make calculations to work out the worst-case scenario possible. For instance, if a rod of 10 cm width is required to fit into a hole, a tight fit is not required. It's possible that the rod itself has errors. As an example, this could mean that it could be as wide as 10.05 cm or as low as 9.95 cm. Logic dictates that the hole needs to be larger than 10.05 cm for a guaranteed fit. This leads to the next possible problem: the hole was produced using a drill size that also has an error. If a drill with a size of 10.05 cm is used, there is always the chance that it will create a hole that is too small.

It is for all these reasons that mathematicians are employed to reduce the chance of waste and potential problems. Indeed, the problem often boils down to one involving statistics and probability; it is the job of quality control to reduce the chance of errors occurring, while maximizing profit and ensuring the equipment actually works as required.

ENGINEERING

Engineers are required to produce and measure important objects for complex design projects. It may take only one defective piece for the whole project to fail. These engineering works will often be integral parts of society. Most of the things taken for granted, such as water, food production, and good health, are direct consequences of engineering projects.

Measuring the Height of Everest

It was during the 1830s that the Great Trigonometrical Survey of The Indian sub-continent was undertaken by William Lambdon. This expedition was one of remarkable human resilience and mathematical application. The aim was to accurately map the huge area, including the Himalayans. Ultimately, they wanted not only the exact location of the many features, but to also evaluate the height above sea level of some of the world's tallest mountains, many of which could not be climbed at that time. How could such a mammoth task be achieved?

Today, it is relatively easy to use trigonometry to estimate how high an object stands. Then, if the position above sea level is known, it takes simple addition to work out the object's actual height compared to Earth's surface. Yet, the main problem for the surveyors in the 1830s was that, although they got within close proximity of the mountains and hence estimated the relative heights, they did not know how high they were above sea level. Indeed they were many hundreds of miles from the nearest ocean.

The solution was relatively simple, though almost unthinkable. Starting at the coast the surveyors would progressively work their way across the vast continent, continually working out heights above sea level of key points on the landscape. This can be referred to in mathematics as an inductive solution. From a simple starting point, repetitions are made until the final solution is found. This method is referred to as triangulation because the key points evaluated formed a massive grid of triangles. In this specific case, this network is often referred to as the great arc.

Eventually, the surveyors arrived deep in the Himalayas and readings from known places were taken; the heights of the mountains were evaluated without even having to climb them! It was during this expedition that a mountain, measured by a man named Waugh, was recorded as reaching the tremendous height of 29,002 feet (recently revised; 8,840 m). That mountain was dubbed Everest, after a man named George Everest who had succeeded Lambdon halfway through the expedition. George Everest never actually saw the mountain.

ARCHAEOLOGY

Archaeology is the study of past cultures, which is important in understanding how society may progress in the future. It can be extremely difficult to explore ancient sites and extract information due to the continual shifting and changing of the surface of the earth. Very few patches of ground are ever left untouched over the years.

While exploring ancient sites, it is important to be able to make accurate representations of the ground. Most items are removed to museums, and so it is important to retain a picture of the ground as originally discovered. A mathematical technique is employed to do so accurately. The distance and depth of items found are measured and recorded, and a map is constructed of the relative positions. Accurate measurements are essential for correct deductions to be made about the history of the site.

ARCHITECTURE

The fact that the buildings we live in will not suddenly fall to the ground is no coincidence. All foundations and structures from reliable architects are built on strict principles of mathematics. They rely upon accurate construction and measurement. With the pressures of deadlines, it is equally important that materials with insufficient accuracy within their measurements are not used.

COMPUTERS

The progression of computers has been quite dramatic. Two of the largest selling points within the computer industry are memory and speed. The speed of a computer is found by measuring the number of calculations that it can perform per second.

BLOOD PRESSURE

When checking the health of a patient, one of the primary factors considered is the strength of the heart, and how it pumps blood throughout the body. Blood pressure measurements reveal how strongly the blood is pumped and other health factors. An accurate measure of blood pressure could ultimately make the difference between life and death.

DOCTORS AND MEDICINE

Doctors are required to perform accurate measurements on a day-to-day basis. This is most evident during surgery where precision may be essential. The

administration of drugs is also subject to precise controls. Accurate amounts of certain ingredients to be prescribed could determine the difference between life and death for the patient.

Doctors also take measurements of patients' temperature. Careful monitoring of this will be used to assess the recovery or deterioration of the patient.

CHEMISTRY

Many of the chemicals used in both daily life and in industry are produced through careful mixture of required substances. Many substances can have lethal consequences if mixed in incorrect doses. This will often require careful measurement of volumes and masses to ensure correct output.

Much of science also depends on a precise measurement of temperature. Many reactions or processes require an optimal temperature. Careful monitoring of temperatures will often be done to keep reactions stable.

NUCLEAR POWER PLANTS

For safety reasons, constant monitoring of the output of power plants is required. If too much heat or dangerous levels of radiation are detected, then action must be taken immediately.

MEASURING TIME

Time drives and motivates much of the activity across the globe. Yet it is only recently that we have been able to measure this phenomenon and to do so consistently. The nature of the modern world and global trade requires the ability to communicate and pass on information at specified times without error along the way.

The ancients used to use the Sun and other celestial objects to measure time. The sundial gave an approximate idea for the time of the day by using the rotation of the Sun to produce a shadow. This shadow then pointed towards a mark/time increment. Unfortunately, the progression of the year changes the apparent motion of the Sun. (Remember, though, that it is due to the change in Earth's orbit around the Sun, not the Sun moving around Earth.) This does not allow for accurate increments such as seconds.

It was Huygens who developed the first pendulum clock. This uses the mathematical principal that the length of a pendulum dictates the frequency with which the pendulum oscillates. Indeed a pendulum of approximately 39 inches will oscillate at a rate of one second. The period of a pendulum is defined to be the time taken for it to do a complete swing to the left, to the right, and back again. These however were not overly accurate, losing many minutes across one day. Yet over time, the accuracy increased.

It was the invention of the quartz clock that allowed much more accurate timekeeping. Quartz crystals vibrate (in a sense, mimicking a pendulum) and this can be utilized in a wristwatch. No two crystals are alike, so there is some natural variance from watch to watch.

THE DEFINITION OF A SECOND

Scientists have long noted that atoms resonate, or vibrate. This can be utilized in the same way as pendulums. Indeed, the second is defined from an atom called cesium. It oscillates at exactly 9,192,631,770 cycles per second.

MEASURING SPEED, SPACE TRAVEL, AND RACING

In a world devoted to transport, it is only natural that speed should be an important measurement. Indeed, the quest for faster and faster transport drives many of the nations on Earth. This is particularly relevant in long-distance travel. The idea of traveling at such speeds that space travel is possible has motivated generations of filmmakers and science fiction authors. Speed is defined to be how far an item goes in a specified time. Units vary greatly, yet the standard unit is meters traveled per second. Once distance and time are measured, then speed can be evaluated by dividing distance by time.

All racing, whether it involves horses or racing cars, will at some stage involve the measuring of speed. Indeed, the most successful sportsperson will be the one who, overall, can go the fastest. This concept of overall speed is often referred to as average speed. For different events, average speed has different meanings.

A sprinter would be faster than a long-distance runner over 100 meters. Yet, over a 10,000-meter race, the converse would almost certainly be true. Average speed gives the true merit of an athlete over the relevant distance. The formula for average speed would be average speed = total distance/total time.

NAVIGATION

The ability to measure angles and distances is an essential ingredient in navigation. It is only through an accurate measurement of such variables that the optimal route can be taken. Most hikers rely upon an advanced knowledge of bearings and distances so that they do not become lost. The same is of course true for any company involved in transportation, most especially those who travel by airplane or ship. There are no roads laid out for

To make a fair race, the tracks must be perfectly spaced. RANDY FARIS/CORBIS.

them to follow, so ability to measure distance and direction of travel are essential.

SPEED OF LIGHT

It is accepted that light travels at a fixed speed through a vacuum. A vacuum is defined as a volume of space containing no matter. Space, once an object has left the atmosphere, is very close to being such. This speed is defined as the speed of light and has a value close to 300,000 kilometers per second.

HOW ASTRONOMERS AND NASA MEASURE DISTANCES IN SPACE

When it comes to the consideration of space travel, problems arise. The distances encountered are so large that if we stick to conventional terrestrial units, the numbers become unmanageable. Distances are therefore expressed as light years. In other words, the distance between two celestial objects is defined to be the time light would take to travel between the two objects.

SPACE TRAVEL AND TIMEKEEPING

The passing of regular time is relied upon and trusted. We do not expect a day to suddenly turn into a year, though psychologically time does not always appear to pass regularly. It has been observed and proven using a branch of mathematics called relativity that, as an object accelerates, so the passing of time slows down for that particular object.

An atomic clock placed on a spaceship will be slightly behind a counterpart left on Earth. If a person could actually travel at speeds approaching the speed of light, they would only age by a small amount, while people on Earth would age normally.

Indeed, it has also been proven mathematically that a rod, if moved at what are classed as relativistic velocities (comparable to the speed of light), will shorten. This is known as the Lorentz contraction. Philosophically, this leads to the question, how accurate are measurements? The simple answer is that, as long as the person and the object are moving at the same speed, then the problem does not arise.

Distance in Three Dimensions

In mathematics it is important to be able to evaluate distance in all dimensions. It is often the case that only the coordinates of two points are known and the distance between them is required. For example, a length of rope needs to be laid across a river so that it is fully taut. There are two trees that have suitable branches to hold the rope on either side. The width of the river is 5 meters. The trees are 3 meters apart widthwise. One of the branches is 1 meter higher than the other. How much rope is required?

The rule is to use an extension of Pythagoras in three dimensions: $a^2 + b^2 = h^2$. An extension to this in three dimensions is: $a^2 + b^2 + c^2 = h^2$. This gives us width, depth, and height. Therefore, $5^2 + 3^2 + 1^2 = h^2 = 35$. Therefore h is just under 6. So at least 6 m of rope is needed to allow for the extra required for tying the knots.

WHY DON'T WE FALL OFF EARTH?

As Isaac Newton sat under a tree, an apple fell off and hit him upon the head. This led to his work on gravity. Gravity is basically the force, or interaction, between Earth and any object. This force varies with each object's mass and also varies as an object moves further away from the surface of Earth.

This variability is not a constant. The reason astronauts on the moon seem to leap effortlessly along is due to the lower force of gravity there. It was essential that NASA was able to measure the gravity on the moon before landing so that they could plan for the circumstances upon arrival.

How is gravity measured on the moon, or indeed anywhere without actually going there first? Luckily, there is an equation that can be used to work it out. This formula relies on knowing the masses of the objects involved and their distance apart.

MEASURING THE SPEED OF GRAVITY

Gravity has the property of speed. Earth rotates about the Sun due to the gravitational pull of the Sun. If the Sun were to suddenly vanish, Earth would continue its orbit until gravity actually catches up with the new situation. The speed of gravity, perhaps unsurprisingly, is the speed of light.

Stars are far away, and we can see them in the sky because their light travels the many light years to meet our retina. It is natural that, after a certain time, most stars end their life often undergo tremendous changes. Were a star to explode and vanish, it could take years for this new reality to be evident from Earth. In fact, some of the stars viewable today may actually have already vanished.

MEASURING MASS

A common theme of modern society is that of weight. A lot of television airplay and books, earning authors millions, are based on losing weight and becoming healthy. Underlying the whole concept of weighing oneself is that of gravity. It is actually due to gravity that an object can actually be weighed.

The weight of an object is defined to be the force that that object exerts due to gravity. Yet these figures are only relevant within Earth's gravity. Interestingly, if a person were to go to the top of a mountain, their measurable weight will actually be less than if they were at sea level. This is simply because gravity decreases the further away an object is from Earth's surface, and so scales measure a lower force from a person's body.

Potential applications

People will continue to take measurements and use them across a vast spectrum of careers, all derived from applications within mathematics. As we move into the future, the tools will become available to increase such measurements to remarkable accuracies on both microscopic and macroscopic levels.

Advancements in medicine and the ability to cure diseases may come from careful measurements within cells and how they interact. The ability to measure, and do so accurately, will drive forward the progress of human society.

Where to Learn More

Periodicals

Muir, Hazel. "First Speed of Gravity Measurement Revealed." *New Scientist.com.*

Web sites

Keay, John. "The Highest Mountain in the World." The Royal Geographical Society. 2003. <http://imagingeverest.rgs.org/Concepts/Virtual_Everest/-288.html> (February 26, 2005).

Medical Mathematics

Overview

Mathematics finds wide applications in medicine and public health. Epidemiology, the scientific discipline that investigates the causes and distribution of disease and that underlies public health practice, relies heavily on mathematical data and analysis. Mathematics is also a critical tool in clinical trials, the cornerstone of medical research supporting modern medical practice, which are used to establish the efficacy and safety of medical treatments. As medical technology and new treatments rely more and more on sophisticated biological modeling and technology, medical professionals will draw increasingly on their knowledge of mathematics and the physical sciences.

There are three major ways in which researchers and practitioners apply mathematics to medicine. The first and perhaps most important is that they must use the mathematics of probability and statistics to make predictions in complex medical situations. The most important example of this is when people try to predict the outcome of illnesses, such as AIDS, cancer, or influenza, in either individual patients or in population groups, given the means that they have to prevent or treat them.

The second important way in which mathematics can be applied to medicine is in modeling biological processes that underlie disease, as in the rate of speed with which a colony of bacteria will grow, the probability of getting disease when the genetics of Mendelian inheritance is known, or the rapidity with which an epidemic will spread given the infectivity and virulence of a pathogen such as a virus. Some of the most commercially important applications of bio-mathematical modeling have been developed for life and health insurance, in the construction of life tables, and in predictive models of health premium increase trend rates.

The third major application of mathematics to medicine lies in using formulas from chemistry and physics in developing and using medical technology. These applications range from using the physics of light refraction in making eyeglasses to predicting the tissue penetration of gamma or alpha radiation in radiation therapy to destroy cancer cells deep inside the body while minimizing damage to other tissues.

While many aspects of medicine, from medical diagnostics to biochemistry, involve complex and subtle applications of mathematics, medical researchers consider epidemiology and its experimental branch, clinical trials, to be the medical discipline for which mathematics is indispensable. Medical research, as furthered by these two disciplines, aims to establish the causes of disease and prove treatment efficacy and safety based on quantitative

(numerical) and logical relationships among observed and recorded data. As such, they comprise the "tip of the iceberg" in the struggle against disease.

The mathematical concepts in epidemiology and clinical research are basic to the mathematics of biology, which is after all a science of complex systems that respond to many influences. Simple or nonstatistical mathematical relationships can certainly be found, as in Mendelian inheritance and bacterial culturing, but these are either the most simple situations or they exist only under ideal laboratory conditions or in medical technology that is, after all, based largely on the physical sciences. This is not to downplay their usefulness or interest, but simply to say that the budding mathematician or scientist interested in medicine has to come to grips with statistical concepts and see how the simple things rapidly get complicated in real life.

Noted British epidemiologist Sir Richard Doll (1912–) has referred to the pervasiveness of epidemiology in modern society. He observed that many people interested in preventing disease have unwittingly practiced epidemiology. He writes, "Epidemiology is the simplest and most direct method of studying the causes of disease in humans, and many major contributions have been made by studies that have demanded nothing more than an ability to count, to think logically and to have an imaginative idea."

Because epidemiology and clinical trials are based on counting and constitute a branch of statistical mathematics in their own right, they require a rather detailed and developed treatment. The presentation of the other major medical mathematics applications will feature explanations of the mathematics that underlie familiar biological phenomena and medical technologies.

Fundamental Mathematical Concepts and Terms

The most basic mathematical concepts in health care are the measures used to discover whether a statistical association exists between various factors and disease. These include rates, proportions, and ratios. Mortality (death) and morbidity (disease) rates are the "raw material" that researchers use in establishing disease causation. Morbidity rates are most usefully expressed in terms of disease incidence (the rate with which population or research sample members contract a disease) and prevalence (the proportion of the group that has a disease over a given period of time).

Beyond these basic mathematical concepts are concepts that measure disease risk. The population at risk is the group of people that could potentially contract a disease, which can range from the entire world population (e.g., at risk for the flu), to a small group of people with a certain gene (e.g., at risk for sickle-cell anemia), to a set of patients that are randomly selected to participate in groups to be compared in a clinical trial featuring alternative treatment modes. Finally, the most basic measure of a population group's risk for a disease is relative risk (the ratio of the prevalence of a disease in one group to the prevalence in another group).

The simplest measure of relative risk is the odds ratio, which is the ratio of the odds that a person in one group has a disease to the odds that a person in a second group has the disease. Odds are a little different from the probability that a person has a disease. One's odds for a disease are the ratio between the number of people that have a disease and the number of people that do not have the disease in a population group. The probability of disease, on the other hand, is the proportion of people that have a disease in a population. When the prevalence of disease is low, disease odds are close to disease probability. For example, if there is a 2%, or 0.02, probability that people in a certain Connecticut county will contract Lyme disease, the odds of contracting the disease will be $2/98 = 0.0204$.

Suppose that the proportion of Americans in a particular ethnic or age group (group 1) with type II diabetes in a given year is estimated from a study sample to be 6.2%, while the proportion in a second ethnic or age group (group 2) is 4.5%. The odds ratio (OR) between the two groups is then: $\text{OR} = (6.2/93.8)/(4.5/95.5) = 0.066/0.047 = 1.403$.

This means that the relative risk of people in group 1 developing diabetes compared to people in group 2 is 1.403, or over 40% higher than that of people in group 2.

The mortality rate is the ratio of the number of deaths in a population, either in total or disease-specific, to the total number of members of that population, and is usually given in terms of a large population denominator, so that the numerator can be expressed as a whole number. Thus in 1982 the number of people in the United States was 231,534,000, the number of deaths from all causes was 1,973,000, and therefore the death rate from all causes of 852.1 per 100,000 per year. That same year there were 1,807 deaths from tuberculosis, yielding a disease-specific mortality rate of 7.8 per million per year.

Assessing disease frequency is more complex because of the factors of time and disease duration. For example, disease prevalence can be assessed at a point in time (point prevalence) or over a period of time (period

prevalence), usually a year (annual prevalence). This is the prevalence that is usually measured in illness surveys that are reported to the public. Researchers can also measure prevalence over an indefinite time period, as in the case of lifetime prevalence. Researchers calculate this time period by asking every person in the study sample whether or not they have ever had the disease, or by checking lifetime health records for everybody in the study sample for the occurrence of the disease, counting the occurrences, and then dividing by the number of people in the population.

The other critical aspect of disease frequency is incidence, which is the number of cases of a disease that occur in a given period of time. Incidence is an extremely critical statistic in describing the course of a fast-moving epidemic, in which medical decision-makers must know how quickly a disease is spreading. The incidence rate is the key to public health planning because it enables officials to understand what the prevalence of a disease is likely to be in the future. Prevalence is mathematically related to the cumulative incidence of a disease over a period of time as well as the expected duration of a disease, which can be a week in the case of the flu or a lifetime in the case of juvenile onset diabetes. Therefore, incidence not only indicates the rate of new disease cases, but is the basis of the rate of change of disease prevalence.

For example, the net period prevalence of cases of disease that have persisted throughout a period of time is the proportion of existing cases at the beginning of that period plus the cumulative incidence during that period of time minus the cases that are cured, self-limited, or that die, all divided by the number of lives in the population at risk. Thus, if there are 300 existing cases, 150 new cases, 40 cures, and 30 deaths in a population of 10,000 in a particular year, the net period (annual) prevalence for that year is $(300 + 150 - 40 - 30) / 10{,}000 = 380/10{,}000 = 0.038$. The net period prevalence for the year in question is therefore nearly 4%.

A crucial statistical concept in medical research is that of the research sample. Except for those studies that have access to disease mortality, incidence, and prevalence rates for the entire population, such as the unique SEER (surveillance, epidemiology and end results) project that tracks all cancers in the United States, most studies use samples of people drawn from the population at risk either randomly or according to certain criteria (e.g., whether or not they have been exposed to a pathogen, whether or not they have had the disease, age, gender, etc.). The size of the research sample is generally determined by the cost of research. The more elaborate and detailed the data collection from the sample participants, the more expensive to run the study.

Medical researchers try to ensure that studying the sample will resemble studying the entire population by making the sample representative of all of the relevant groups in the population, and that everyone in the relevant population groups should have an equal chance of getting selected into the sample. Otherwise the sample will be biased, and studying it will prove misleading about the population in general.

The most powerful mathematical tool in medicine is the use of statistics to discover associations between death and disease in populations and various factors, including environmental (e.g., pollution), demographic (age and gender), biological (e.g., body mass index, or BMI), social (e.g., educational level), and behavioral (e.g., tobacco smoking, diet, or type of medical treatment), that could be implicated in causing disease.

Familiarity with basic concepts of probability and statistics is essential in understanding health care and clinical research and is one of the most useful types of knowledge that one can acquire, not just in medicine, but also in business, politics, and such mundane problems as interpreting weather forecasts.

A statistical association takes into account the role of chance. Researchers compare disease rates for two or more population groups that vary in their environmental, genetic, pathogen exposure, or behavioral characteristics, and observe whether a particular group characteristic is associated with a difference in rates that is unlikely to have occurred by chance alone.

How can scientists tell whether a pattern of disease is unlikely to have occurred by chance? Intuition plays a role, as when the frequency of disease in a particular population group, geographic area, or ecosystem is dramatically out of line with frequencies in other groups or settings. To confirm the investigator's hunches that some kind of statistical pattern in disease distribution is emerging, researchers use probability distributions.

Probability distributions are natural arrays of the probability of events that occur everywhere in nature. For example, the probability distribution observed when one flips a coin is called the binomial distribution, so-called because there are only two outcomes: heads or tails, yes or no, on or off, 1 or 0 (in binary computer language). In the binomial distribution, the expected frequency of heads and tails is 50/50, and after a sufficiently long series of coin flips or trials, this is indeed very close to the proportions of heads and tails that will be observed. In medical research, outcomes are also often binary, i.e., disease is

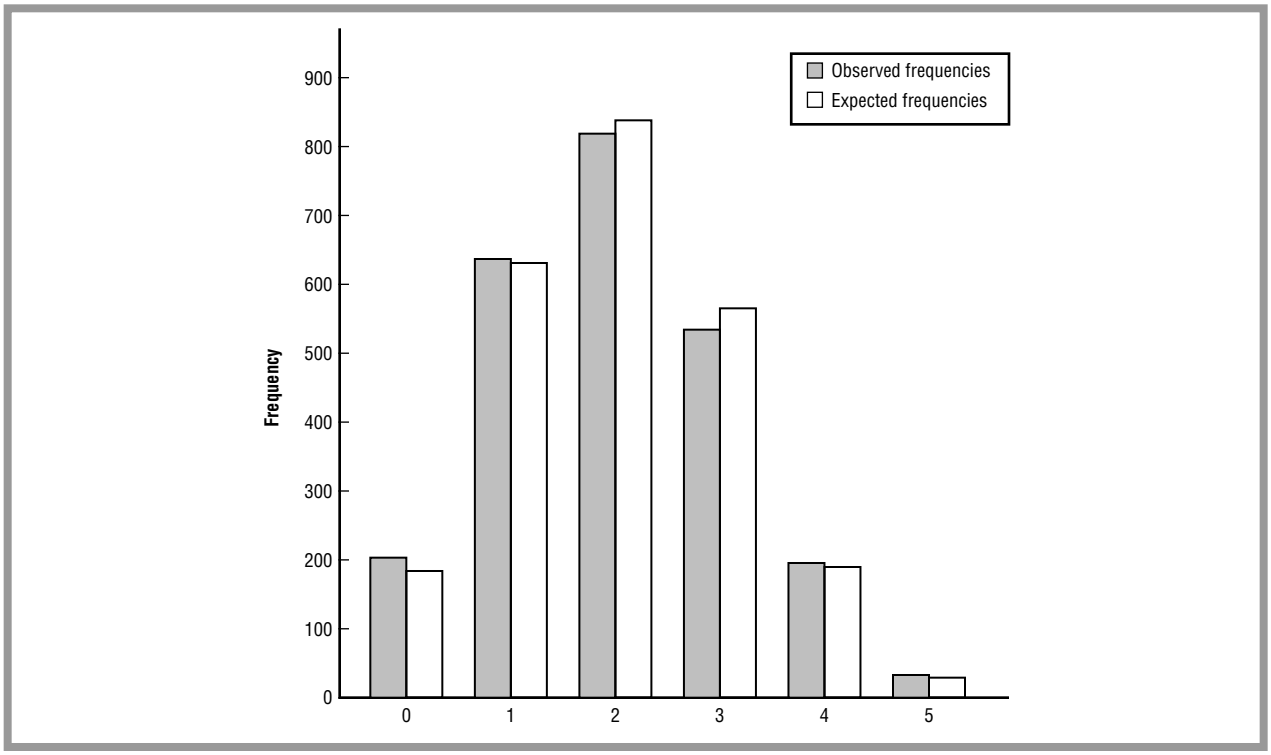


Figure 1: Binomial distribution.

present or absent, exposure to a virus is present or absent, the patient is cured or not, the patient survives or not.

However, people almost never see exactly 50/50, and the shorter the series of coin flips, the bigger the departure from 50/50 will likely be observed. The binomial probability distribution does all of this coin-flipping work for people. It shows that 50/50 is the expected odds when nothing but chance is involved, but it also shows that people can expect departures from 50/50 and how often these departures will happen over the long run. For example, a 60/40 odds of heads and tails is very unlikely if there are 30 coin tosses (18 heads, 12 tails), but much more likely if one does only five coin tosses (e.g., three heads, two tails). Therefore, statistics books show binomial distribution tables by the number of trials, starting with $n = 5$, and going up to $n = 25$. The binomial distribution for ten trials is a "stepwise," or discrete distribution, because the probabilities of various proportions jump from one value to another in the distribution. As the number of trials gets larger, these jumps get smaller and the binomial distribution begins to look smoother. Figure 1 provides an illustration of how actual and expected outcomes might differ under the binomial distribution.

Beyond $n = 30$, the binomial distribution becomes very cumbersome to use. Researchers employ the normal distribution to describe the probability of random events in larger numbers of trials. The binomial distribution is said to approach the normal distribution as the number of trials or measurements of a phenomenon get higher. The normal distribution is represented by a smooth bell curve. Both the binomial and the normal distributions share in common that the expected odds (based on the mean or average probability of 0.5) of "on-off" or binary trial outcomes is 50/50 and the probabilities of departures from 50/50 decrease symmetrically (i.e., the probability of 60/40 is the same as that of 40/60). Figure 2 provides an illustration of the normal distribution, along with its cumulative S-curve form that can be used to show how random occurrences might mount up over time.

In Figure 2, the expected (most frequent) or mean value of the normal distribution, which could be the average height, weight, or body mass index of a population group, is denoted by the Greek letter μ, while the standard deviation from the mean is denoted by the Greek letter σ. Almost 70% of the population will have a measurement that is within one standard deviation

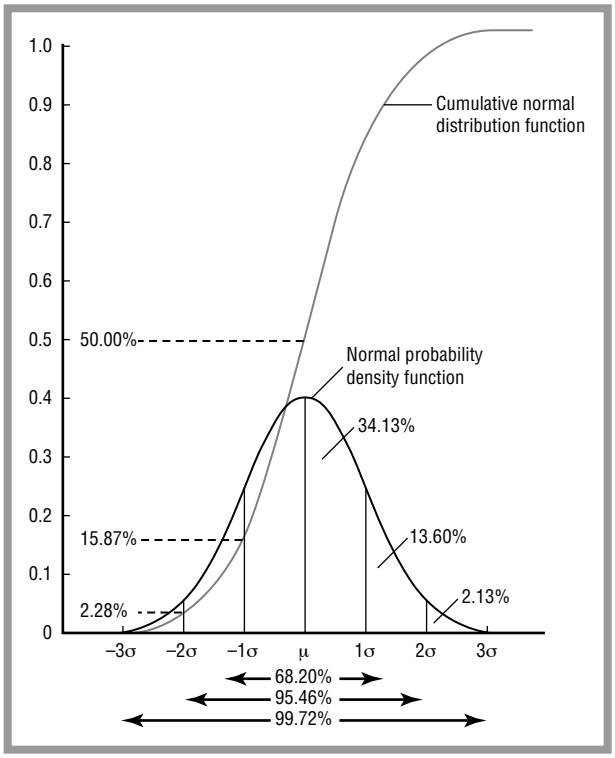


Figure 2: Population height and weight.

from the mean; on the other hand, only about 5% will have a measurement that is more than two standard deviations from the mean. The low probability of such measurements has led medical researchers and statisticians to posit approximately two standard deviations as the cutoff point beyond which they consider an occurrence to be significantly different from average because there is only a one in 20 chance of its having occurred simply by chance.

The steepness with which the probability of the odds decreases as one continues with trials determines the width or variance of the probability distribution. Variance can be measured in standardized units, called standard deviations. The further out toward the low probability tails of the distribution the results of a series of trials are, the more standard deviations from the mean, and the more remarkable they are from the investigator's standpoint. If the outcome of a series of trials is more than two standard deviations from the mean outcome, it will have a probability of 0.05 or one chance in 20. This is the cutoff, called the alpha (α) level beyond which researchers usually judge that the outcome of a series of trials could not have occurred by chance alone. At that point they begin to consider that one or more factors are causing the observed pattern. For example, if the frequency pattern of disease is similar to the frequencies of age, income, ethnic groups, or other features of population groups, it is usually a good bet that these characteristics of people are somehow implicated in causing the disease, either directly or indirectly.

The normal distribution helps disease investigators decide whether a set of odds (e.g., 10/90) or a probability of 10% of contracting a disease in a subgroup of people that behave differently from the norm (e.g., alcoholics) is such a large deviation (usually, more than two standard deviations) from the expected frequency that the departure exceeds the alpha level of a probability of 0.05. This deviation would be considered to be statistically significant. In this case, a researcher would want to further investigate the effect of the behavioral difference. Whether or not a particular proportion or disease prevalence in a subgroup is statistically significant depends on both the difference from the population prevalence as well as the number of people studied in the research sample.

Real-life Applications

VALUE OF DIAGNOSTIC TESTS

Screening a community using relatively simple diagnostic tests is one of the most powerful tools that health care professionals and public health authorities have in preventing disease. Familiar examples of screening include HIV testing to help prevent AIDS, cholesterol testing to help prevent heart disease, mammography to help prevent breast cancer, and blood pressure testing to help prevent stroke. In undertaking a screening program, authorities must always judge whether the benefits of preventing the illness in question outweigh the costs and the number of cases that have been mistakenly identified, called false positives.

Every diagnostic or screening test has four basic mathematical characteristics: sensitivity (the proportion of identified cases that are true cases), specificity (the proportion of identified non-cases that are true non-cases), positive predictive value (PV^+, the probability of a positive diagnosis if the case is positive), and negative predictive value (PV^-, the probability of a negative diagnosis if the case is negative). These values are calculated as follows. Let a = the number of identified cases that are real cases of the disease (true positives), b = the number of identified cases that are not real cases (false positives), c = the number of true cases that were not identified by the test (false negatives), and d = the number of individuals identified as non-cases that were true non-cases (true negatives). Thus, the number of true cases is a + c,

A researcher collects blood from a "sentinel" chicken from an area being monitored for the West Nile virus. FADEK TIMOTHY/CORBIS SYGMA.

the number of true non-cases is b + d, and the total number of cases is a + b + c + d. The four test characteristics or parameters are thus Sensitivity = a/a + b; Specificity = d/b + d; PV^+ = a/a + b; PV^- = d/c + d. These concepts are illustrated in Table 1 for a mammography screening study of nearly 65,000 women for breast cancer.

Calculating the four parameters of the screening test yields: Sensitivity = 132 / 177 = 74.6%; Specificity = 63,650 / 64, 633 = 98.5%; PV^+ = 132 / 1,115 = 11.8%; PV^- = 63,650 / 63,695 = 99.9%.

These parameters, especially the ability of the test to identify true negatives, make mammography a valuable prevention tool. However, the usefulness of the test is proportional to the disease prevalence. In this case, the disease prevalence is very low: (a + c)/(b + d) = 177/64,683 ≈ 0.003, and the positive predictive value is less than 12%. In other words, the actual cancer cases identified are a small minority of all of the positive cases. As the prevalence of breast cancer rises, as in older women, the proportion of actual cases rises. This makes the test much more cost effective when used on women over the age of 50 because the proportion of women that undergo expensive biopsies that do not confirm the mammography results is much lower than if mammography was administered to younger women or all women.

CALCULATION OF BODY MASS INDEX (BMI)

The body mass index (BMI) is often used as a measure of obesity, and is a biological characteristic of individuals that is strongly implicated in the development or etiology of a number of serious diseases, including diabetes and heart disease. The BMI is a person's weight, divided by his or her height squared: BMI = weight/height^2. For example, if a man is 1.8 m tall and weighs 85 kg, his body mass index is: 85 kg^2/1.8 m = 26.2. For BMIs over 26, the risk of diabetes and coronary artery disease is elevated, according to epidemiological studies. However, a more recent study has shown that stomach girth is more strongly related to diabetes risk than BMI itself, and BMI may not be a reliable estimator of disease risk for athletic people with more lean muscle mass than average.

STANDARD DEVIATION AND VARIANCE FOR USE IN HEIGHT AND WEIGHT CHARTS

Concepts of variance and the standard deviation are often depicted in population height and weight charts.

Suppose that the average height of males in a population is 1.9 meters. Investigators usually want to know more than just the average height. They might also like to know the frequency of other heights (1.8 m, 2.0 m, etc.). By studying a large sample, say 2,000 men from the population, they can directly measure the men's heights and calculate a convenient number called the sample's standard deviation, by which they could describe how close or how far away from the average height men in this population tend to be. To get this convenient number, the researchers simply take the average difference from the mean height. To do this, they would first sum up all of these differences or deviations from average, and then divide by the number of men measured. To use a simple example, suppose five men from the population are measured and their heights are 1.8 m, 1.75 m, 2.01 m, 2.0 m, and 1.95 m. The average or mean height of this small sample in meters = (1.8 + 1.75 + 2.01 + 2.0 + 1.95)/5 = 1.902. The difference of each man's height from the average height of the sample, or the deviation from average. The sample standard deviation is simply the average

Counting calories is a practice of real-life mathematics that can have a dramatic impact on health. A collection of menu items from opposite ends of the calorie spectrum including a vanilla shake from McDonald's (1,100 calories); a Cuban Panini sandwich from Ruby Tuesday's (1,164 calories), and a six-inch ham sub, left, from Subway (290 calories). All the information for these items is readily available at the restaurants that serve them. AP/WIDE WORLD PHOTOS. REPRODUCED BY PERMISSION.

of the deviations from the mean. The deviations are $1.8 - 1.902 = -0.102$, $1.75 - 1.902 = -0.152$, $2.01 - 1.902 = 0.108$, $2.0 - 1.902 = 0.008$, and $1.95 - 1.902 = 0.048$. Therefore, the average deviation for the sample is $(-1.02 - 0.152 + 0.108 + 0.008 + 0.048)/5 = -0.2016$ m.

However, this is a negative number that is not appropriate to use because the standard deviation is supposed to be a directionless unit, as is an inch, and because the average of all of the average deviations will not add up to the population average deviation. To get the sample standard deviation to always be positive, no matter which sample of individuals that is selected to be measured, and to ensure that it is a good estimator of the population average deviation, researchers go through additional steps. They sum up the squared deviations, calculate the average squared deviation (mean squared deviation), and take the square root of the sum of the squared deviations (the root mean squared deviation or RMS deviation). They then add a correction factor of –1 in the denominator.

So the sample standard deviation in the example is

$$s = \sqrt{\frac{(-.102)^2 + (-.152)^2 + (.108)^2 + (.008)^2 + (.048)^2}{4}} \cong 0.109$$

Note that the sample average of 1.902 m happens in this sample to be close to the known population average, denoted as μ, of 1.9 m. The sample standard deviations might or might not be close to the population standard deviation, denoted as σ. Regardless, the sample average and standard deviation are both called estimators of the population average and standard deviation. In order for any given sample average or standard deviation to be considered to be an accurate estimator for the population average and standard deviation, a small correction factor is applied to these estimators to take into account that a sample has already been drawn, which puts a small constraint (eliminates a degree of freedom) on the estimation of μ and σ for the population. This is done so that after many samples are examined, the mean of all the sample means and the average of all of the sample standard deviations approaches the true population mean and standard deviation.

GENETIC RISK FACTORS: THE INHERITANCE OF DISEASE

Nearly all diseases have both genetic (heritable) and environmental causes. For example, people of Northern European ancestry have a higher incidence of skin cancer from sun exposure in childhood than do people of Southern European or African ancestry. In this case, Northern Europeans' lack of skin pigment (melanin) is the heritable part, and their exposure to the sun to the point of burning, especially during childhood, is the environmental part. The proportion of risk due to inheritance and the proportion due to the environment are very difficult to figure out. One way is to look at twins who have the same genetic background, and see how often various environmental differences that they have experienced have resulted in different disease outcomes.

However, there is a large class of strictly genetic diseases for which predictions are fairly simple. These are diseases that involve dominant and recessive genes. Many genes have alternative genotypes or variants, most of which are harmful or deleterious. Each person receives

Real-life sensitivity and specificity in cancer screening

Screening test (mammography)	Cancer confirmed	Cancer not confirmed	Total
Positive	a = 132	b = 983	**a + b = 1,115**
Negative	c = 45	d = 63,650	**c + d = 63,695**
Total	**a + c = 177**	**b + d = 64,683**	**a + b + c + d = 64,810**

Table 1.

one of these gene variants from each parent, so he or she has two variants for each gene that vie for expression as one grows up. People express dominant genes when the variant contributed by one parent overrides expression of the other parent's variant (or when both parents have the same dominant variant). Some of these variants make the fetus a "non-starter," and result in miscarriage or spontaneous abortion. Other variants do not prevent birth and may not express disease until middle age. In writing about simple Mendelian inheritance, geneticists can use the notation AA to denote homozygous dominant (usually homozygous normal), Aa to denote heterozygous recessive, and aa to denote homozygous recessive.

One tragic example is that of Huntington's disease due to a dominant gene variant, in which the nervous system deteriorates catastrophically at some point after the age of 35. In this case, the offspring can have one dominant gene (Huntington's) and one normal gene (heterozygous dominant), or else can be homozygous dominant (both parents had Huntington's disease, but had offspring before they started to develop symptoms). Because Huntington's disease is caused by a dominant gene, the probability of the offspring developing the disease is 100%.

When a disease is due to a recessive gene allele or variant, one in which the normal gene is expressed in the parents, the probability of inheriting the disease is slightly more complicated. Suppose that two parents are heterozygous recessive (both are Aa). The pool of variants contributed by both parents that can be distributed to the offspring, two at a time, are thus A, A, a, and a. Each of the four gene variant combinations (AA, Aa, aA, aa) has a 25% chance of being passed on to an offspring. Three of these combinations produce a normal offspring and one produces a diseased offspring, so the probability of contracting the recessive disease is 25% under the circumstances.

In probability theory, the probability of two events occurring together is the product of the probability of each of the two events occurring separately. So, for example, the probability of the offspring getting AA is $\frac{1}{2} \times \frac{1}{2} = \frac{1}{4}$ (because half of the variants are A), the probability of getting Aa is $2 \times \frac{1}{4} = \frac{1}{2}$ (because there are two ways of becoming heterozygous), and the probability of getting aa is $\frac{1}{4}$ (because half of the variants are a). Only one of these combinations produces the recessive phenotype that expresses disease.

Therefore, if each parent is heterozygous recessive (Aa), the offspring has a 50% chance of receiving aa and getting the disease. If only one parent is heterozygous normal (Aa) and the other is homozygous recessive (aa), and the disease has not been devastatingly expressed before childbearing age, then the offspring will have a 75% chance of inheriting the disease. Finally, if both parents are homozygous recessive, then the offspring will have a 100% chance of developing the disease.

Some diseases show a gradation between homozygous normal, heterozygous recessive, and homozygous recessive. An example is sickle-cell anemia, a blood disease characterized by sickle-shaped red blood cells that do not efficiently convey oxygen from the lungs to the body, found most frequently in African populations living in areas infested with malaria carried by the tsetse fly. Let AA stand for homozygous for the normal, dominant genotype, Aa for the heterozygous recessive genotype, and aa for the homozygous recessive sickle-cell genotype. It turns out that people living in these areas with the normal genotype are vulnerable to malaria, while people carrying the homozygous recessive genotype develop sickle-cell anemia and die prematurely. However, the heterozygous individuals are resistant to malaria and rarely develop sickle-cell anemia; therefore, they actually have an advantage in surviving or staying healthy long enough to bear children in these regions. Even though the sickle-cell variant leads to devastating disease that prevents an individual from living long enough to reproduce, the population in the tsetse fly regions gets a great benefit from having this variant in the gene pool. Anthropologists cite the distribution of sickle-cell anemia as evidence of how environmental conditions influence the gene pool in a population and result in the evolution of human traits.

The inheritance of disease becomes more and more complicated as the number of genes involved increase. At

How Simple Counting has Come to be the Basis of Clinical Research

The first thinker known to consider the fundamental concepts of disease causation was none other than the ancient Greek physician Hippocrates (460–377 B.C.), when he wrote that medical thinkers should consider the climate and seasons, the air, the water that people use, the soil and people's eating, drinking, and exercise habits in a region. Subsequently, until recent times, these causes of diseases were often considered but not quantitatively measured. In 1662 John Graunt, a London haberdasher, published an analysis of the weekly reports of births and deaths in London, the first statistical description of population disease patterns. Among his findings he noted that men had a higher death rate than women, a high infant mortality rate, and seasonal variations in mortality. Graunt's study, with its meticulous counting and disease pattern description, set the foundation for modern public health practice.

Graunt's data collection and analytical methodology was furthered by the physician William Farr, who assumed responsibility for medical statistics for England and Wales in 1839 and set up a system for the routine collection of the numbers and causes of deaths. In analyzing statistical relationships between disease and such circumstances as marital status, occupations such as mining and working with earthenware, elevation above sea level, and imprisonment, he addressed many of the basic methodological issues that contemporary epidemiologists deal with. These include defining populations at risk for disease and the relative disease risk between population groups, and considering whether associations between disease and the factors mentioned above might be caused by other factors, such as age, length of exposure to a condition, or overall health.

A generation later, public health research came into its own as a practical tool when another British physician, John Snow, tested the hypothesis that a cholera epidemic in London was being transmitted by contaminated water. By examining death rates from cholera, he realized that they were significantly higher in areas supplied with water by the Lambeth and the Southwark and Vauxhall companies, which drew their water from a part of the Thames River that was grossly polluted with sewage. When the Lambeth Company changed the location of its water source to another part of the river that was relatively less polluted, rates of cholera in the areas served by that company declined, while no change occurred among the areas served by the Southwark and Vauxhall. Areas of London served by both companies experienced a cholera death rate that was intermediate between the death rates in the areas supplied by just one of the companies. In recognizing the grand but simple natural experiment posed by the change in the Lambeth Company water source, Snow was able to make a uniquely valuable contribution to epidemiology and public health practice.

After Snow's seminal work, epidemiologists have come to include many chronic diseases with complex and often still unknown causal agents; the methods of epidemiology have become similarly complex. Today researchers use genetics, molecular biology, and microbiology as investigative tools, and the statistical methods used to establish relative disease risk draw on the most advanced statistical techniques available.

Yet reliance on meticulous counting and categorizing of cases and the imperative to think logically and avoid the pitfalls in mathematical relationships in medical data remain at the heart of all of the research used to prove that medical treatments are safe and effective. No matter how high technology, such as genetic engineering or molecular biology, changes the investigations of basic medical research, the diagnostic tools and treatments that biochemists or geneticists propose must still be adjudicated through a simple series of activities that comprise clinical trials: random assignments of treatments to groups of patients being compared to one another, counting the diagnostic or treatment outcomes, and performing a simple statistical test to see whether or not any differences in the outcomes for the groups could have occurred just by chance, or whether the newfangled treatment really works. Many hundreds of millions of dollars have been invested by governments and pharmaceutical companies into ultra-high technology treatments only to have a simple clinical trial show that they are no better than placebo. This makes it advisable to keep from getting carried away by the glamour of exotic science and technologies when it comes to medicine until the chickens, so to speak, have all been counted.

a certain point, it is difficult to determine just how many genes might be involved in a disease—perhaps hundreds of genes contribute to risk. At that point, it is more useful to think of disease inheritance as being statistical or quantitative, although new research into the human genome holds promise in revealing how information about large numbers of genes can contribute to disease prognosis and treatment.

CLINICAL TRIALS

Clinical trials constitute the pinnacle of Western medicine's achievement in applying science to improve human life. Many professionals find trial work very exciting, even though it is difficult, exacting, and requires great patience as they anxiously await the outcomes of trials, often over periods of years. It is important that the sense of drama and grandeur of the achievements of the trials should be passed along to young people interested in medicine. There are four important clinical trials currently in the works, the results of which affect the lives and survival of hundreds of thousands, even millions, of people, young and old.

The first trial was a rigorous test of the effectiveness of condoms in HIV/AIDS prevention. This was a unique experiment reported in 1994 in the *New England Journal of Medicine* that appears to have been under-reported in the popular press. Considering the prestige of the Journal and its rigorous peer-review process, it is possible that many lives could be saved by the broader dissemination of this kind of scientific result. The remaining three trials are a sequence of clinical research that have had a profound impact on the standard of breast cancer treatment, and which have resulted in greatly increased survival. In all of these trials, the key mathematical concept is that of the survival function, often represented by the Kaplan-Meier survival curve, shown in Figure 4 below.

Clinical trial 1 was a longitudinal study of human immunodeficiency virus (HIV) transmission by heterosexual partners Although in the United States and Western Europe the transmission of AIDS has been largely within certain high-risk groups, including drug users and homosexual males, worldwide the predominant mode of HIV transmission is heterosexual intercourse. The effectiveness of condoms to prevent it is generally acknowledged, but even after more than 25 years of the growth of the epidemic, many people remain ignorant of the scientific support for the condom's preventive value.

A group of European scientists conducted a prospective study of HIV negative subjects that had no risk factor for AIDS other than having a stable heterosexual relationship with an HIV infected partner. A sample of 304 HIV negative subjects (196 women and 108 men) was followed for an average of 20 months. During the trial, 130 couples (42.8%) ended sexual relations, usually due to the illness or death of the HIV-infected partner. Of the remaining 256 couples that continued having exclusive sexual relationships, 124 couples (48.4%) consistently used condoms. None of the seronegative partners among these couples became infected with HIV. On the other hand, among the 121 couples that inconsistently used condoms, the seroconversion rate was 4.8 per 100 person-years (95% confidence interval, 2.5–8.4). This means that inconsistent condom-using couples would experience infection of the originally uninfected partner between 2.5 and 8.4 times for every 100 person-years (obtained by multiplying the number of couples by the number of years they were together during the trial), and the researchers were confident that in 95 times out of a 100 trials of this type, the seroconversion rate would lie in this interval. The remaining 11 couples refused to answer questions about condom use. HIV transmission risk increased among the inconsistent users only when infected partners were in the advanced stages of disease ($p < 0.02$) and when the HIV negative partners had genital infections ($p < 0.04$).

Because none of the seronegative partners among the consistent condom-using couples became infected, this trial presents extremely powerful evidence of the effectiveness of condom use in preventing AIDS. On the other hand, there appear to be several main reasons why some of the couples did not use condoms consistently. Therefore, the main issue in the journal article shifts from the question of whether or not condoms prevent HIV infection—they clearly do—to the issue of why so many couples do not use condoms in view of the obvious risk. Couples with infected partners that got their infection through drug use were much less likely to use condoms than when the seropositive partner got infected through sexual relations. Couples with more seriously ill partners at the beginning of the study were significantly more likely to use condoms consistently. Finally, the couples who had been together longer before the start of the trial were positively associated with condom use.

Clinical trial 2 investigated the survival value of dense-dose ACT with immune support versus ACT given in three-week cycles Breast cancer is particularly devastating because a large proportion of cases are among young and middle-aged women in the prime of life. The majority of cases are under the age of 65 and the most aggressive cases occur in women under 50. The very most aggressive cases occur in women in their 20s, 30s, and 40s. The development of the National Cancer Care Network (NCCN) guidelines for treating breast cancer is the result

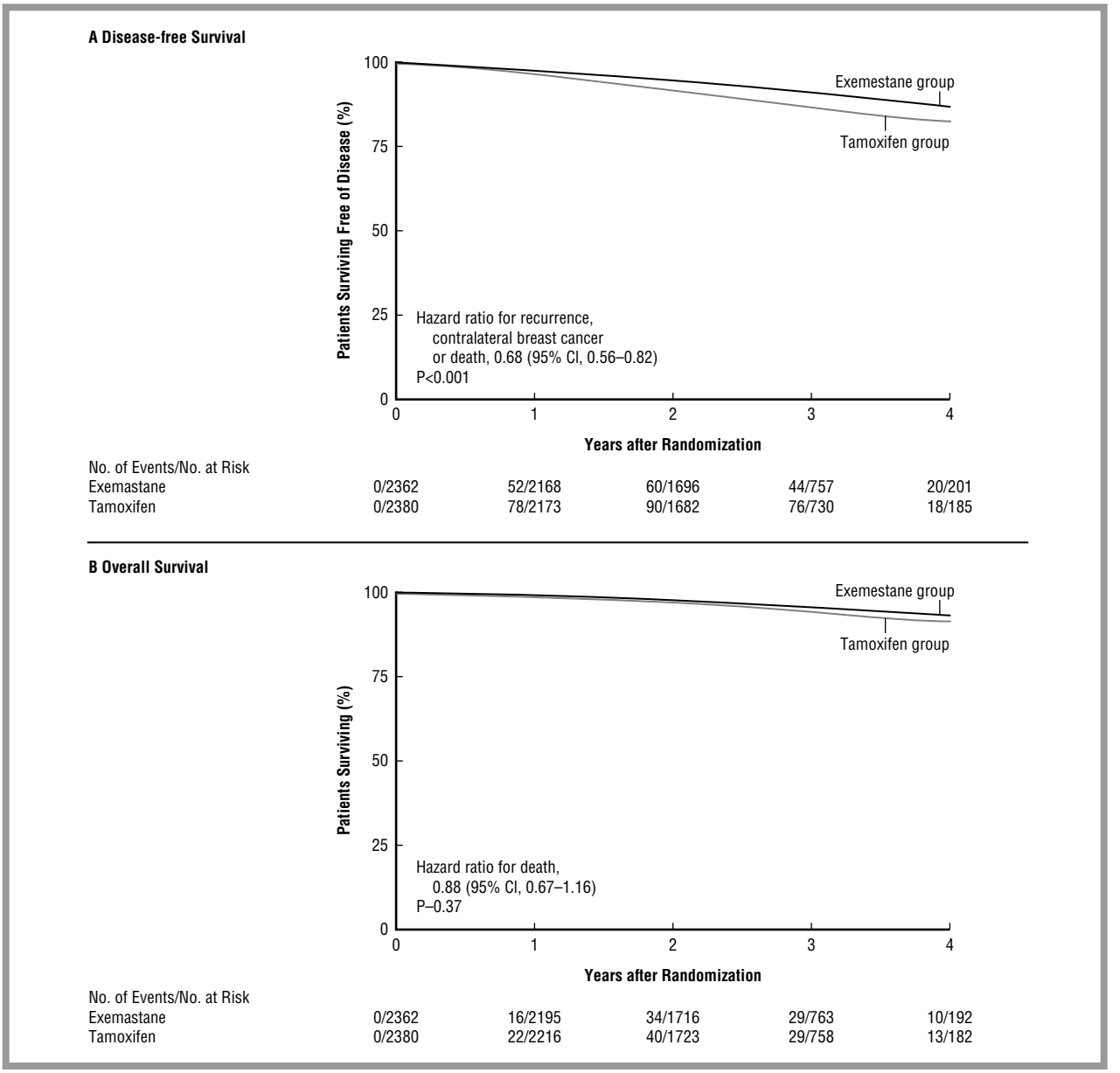


Figure 4: Cancer survival data.

of an accumulation of clinical trial evidence over many years. At each stage of the NCCN treatment algorithm, the clinician must make a treatment decision based on the results of cancer staging and the evidence for long-term (generally five-year) survival rates from clinical trials.

A treatment program currently recommended in the guidelines for breast cancer that is first diagnosed is that the tumor is excised in a lumpectomy, along with any lymph nodes found to contain tumor cells. Some additional nodes are usually removed in determining how far the tumor has spread into the lymphatic system. The tumor is tested to see whether it is stimulated by estrogen or progesterone. If so, the patient is then given chemotherapy with a combination of doxorubicin (Adriamycin) plus cyclophosphamide (AC) followed by paclitaxel (Taxol, or T) (the ACT regimen). In the original protocol, doctors administered eight chemotherapy infusion cycles (four AC and four T) every three weeks to give the patient's immune system time to recover. The patient then receives radiation therapy for six weeks. After radiation, the patient receives either Tamoxifen or an aromatase inhibitor for years as secondary preventive treatment.

Oncologists wondered whether compressing the three-week cycles to two weeks (dense dosing) while supporting the immune system with filgrastim, a white cell growth factor, would further improve survival. They speculated that dense dosing would reduce the opportunity for cancer cells to recover from the previous cycle and continue to multiply. Filgrastim was used between cycles because a patient's white cell count usually takes about three weeks to recover spontaneously from a chemotherapy infusion, and this immune stimulant has been shown to shorten recovery time.

The researchers randomized 2,005 patients into four treatment arms: 1) A-C-T for 36 weeks, 2) A-C-T for 24 weeks, 3) AC-T for 24 weeks, and 4) AC-T for 16 weeks. The patients in the dense dose arms (2 and 4) received filgrastim. These patients were found to be less prone to infection than the patients in the other arms (1 and 3).

After 36 months of follow-up, the primary endpoint of disease-free survival favored the dense dose arms with a 26% reduction in the risk of recurrence. The probability of this result by chance alone was only 0.01 ($p = 0.01$), a result that the investigators called exciting and encouraging. Four-year disease-free survival was 82% in the dense-dose arms versus 75% for the other arms. Results were also impressive for the secondary endpoint of overall survival. Patients treated with dense-dose therapy had a mortality rate 31% lower than those treated with conventional therapy ($p = 0.013$). They had an overall four-year survival rate of 92% compared with 90% for conventional therapy. No significant difference in the primary or secondary endpoints was observed between the A-C-T patients versus the AC-T patients: only dense dosing made a difference. The benefit of the AC-T regimen was that patients were able to finish their therapy eight weeks earlier, a significant gain in quality of life when one is a cancer patient.

One of the salient mathematical features of this trial is that it had enough patients (2,005) to be powered to detect such a small difference (2%) in overall survival rate. Many trials with fewer than 400 patients in total are not powered to detect differences with such precision. Had this difference been observed in a smaller trial, the survival difference might not have been statistically significant.

Clinical trial 3 studied the treatment of patients over 50 with radiation and tamoxifen versus tamoxifen alone. Some oncologists have speculated that women over 50 may not get additional benefit by receiving radiation therapy after surgery and chemotherapy. A group of Canadian researchers set up a clinical trial to test this hypothesis that ran between 1992 and 2000 involving women 50 years or older with early stage node-negative breast cancer with tumors 5 cm in diameter or less. A sample of 769 women was randomized into two treatment arms: 1) 386 women received breast irradiation plus tamoxifen, and 2) 383 women received tamoxifen alone. They were followed up for a median of 5.6 years.

The local recurrence rate (reappearance of the tumor in the same breast) was 7.7% in the tamoxifen group and 0.6% in the tamoxifen plus radiation group. Analysis of the results produced a hazard ratio of 8.3 with a 95% confidence interval of [3.3, 21.2]. This means that women in the tamoxifen group were more than eight times as likely to have local tumor recurrences than the group that received irradiation, and the researchers were confident that in 95 times out of a 100 trials of this type, the hazard ratio would at least be over three times as great and as much as 21.2 times as great, given the role of random chance fluctuations. The probability of this result was that it could occur by chance alone only once in a 1,000 trials ($p < 0.001$).

As mentioned above, clinical trials are the interventional or experimental application of epidemiology and constitute a unique branch of statistical mathematics. Statisticians that are specialists in such studies are called trialists. Clinical trial shows how the rigorous pursuit of clinical trial theory can result in some interesting and perplexing conundrums in the practice of medicine.

In this trial, they studied the secondary prevention effectiveness of tamoxifen versus Exemestane. For the past 20 years, the drug tamoxifen (Nolvadex) has been the standard treatment to prevent recurrence of breast cancer after a patient has received surgery, chemotherapy, and radiation. It acts by blocking the stimulatory action of estrogen (the female hormone estrogen can stimulate tumor growth) by binding to the estrogen receptors on breast tumor cells (the drug is an estrogen imitator or agonist). The impact of tamoxifen on breast cancer recurrence (a 47% decrease) and long-term survival (a 26% increase) could hardly be more striking, and the life-saving benefit to hundreds of thousands of women has been one of the greatest success stories in the history of cancer treatment. One of the limitations of tamoxifen, however, is that after five years patients generally receive no benefit from further treatment, although the drug is considered to have a "carryover effect" that continues for an indefinite time after treatment ceases.

Nevertheless, over the past several years a new class of endocrine therapy drugs called aromatase inhibitors (AIs) that have a different mechanism or mode of action from that of tamoxifen have emerged. AIs have an even more complete anti-estrogen effect than tamoxifen, and

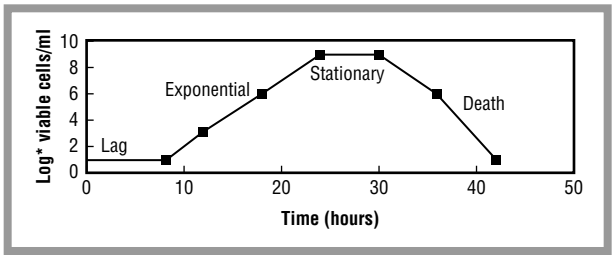


Figure 5: Bacterial growth curve for viable (living) cells.

showed promise as a treatment that some patients could use after their tumors had developed resistance to tamoxifen. As recently as 2002 the medical information company WebMD published an Internet article reporting that some oncologists still preferred the tried-and-true tamoxifen to the newcomer AIs despite mounting evidence of their effectiveness.

However, the development of new "third generation" aromatase inhibitors has spurred new clinical trials that now make it likely that doctors will prescribe an AI for new breast cancer cases that have the most common patient profile (stages I–IIIa, estrogen sensitive) or for patients that have received tamoxifen for 2–5 years. A very large clinical trial reported in 2004 addressed switching from tamoxifen to an AI. A large group of 4,742 postmenopausal patients over age 55 with primary (nonmetastatic) breast cancer that had been using tamoxifen for 2–3 years was recruited into the trial between February 1998 and February 2003. About half (2,362) were randomly assigned (randomized) into the exemestane group and the remainder (2,380) were randomized into the tamoxifen group (the group continuing their tamoxifen therapy). Disease-free survival, defined as the time from the start of the trial to the recurrence of the primary tumor or occurrence of a contralateral (opposite breast) or a metastatic tumor, was the primary trial endpoint.

In all, 449 first events (new tumors) were recorded, 266 in the tamoxifen group and 183 in the exemestane group, by June 30, 2003. This large excess of events in the tamoxifen group was highly statistically significant ($p < 0.0004$, known as the O'Brien-Fleming stopping boundary), and the trial's data and safety-monitoring committee, a necessary component of all clinical trials, recommended an early halt to the trial. Trial oversight committees always recommend an early trial ending when preliminary results are so statistically significant that continuing the trial would be unethical. This is because continuation would put the lives of patients in one of the trial arms at risk because they were not receiving medication that had already shown clear superiority.

The unadjusted hazard ratio for the exemestane group compared to the tamoxifen group was 0.62 (95% confidence interval 0.56–0.82, $p < 0.00005$, corresponding to an absolute benefit of 4.7%). Disease-free survival in the exemestane group was 91.5% (95% confidence interval 90.0–92.7%) versus 86.8% for the tamoxifen group (95% confidence interval 85.1–88.3%). The 95% confidence interval around the average disease-free survival rate for each group is a band of two standard errors (related to the standard deviation) on each side. If these bands do not overlap, as these did not, the difference in disease-free survival for the two groups is statistically significant.

The advantage of exemestane was even greater when deaths due to causes other than breast cancer were censored (not considered in the statistical analysis) in the results. One important ancillary result, however, was that at the point the trial was discontinued; there was no statistically significant difference in overall survival between the two groups. This prompted an editorial in the *New England Journal of Medicine* that raised concern that many important clinical questions that might have been answered had the trial continued, such as whether tamoxifen has other benefits, for instance osteoporosis and cardiovascular disease prevention effects, in breast cancer patients, now could not be and perhaps might never be addressed.

RATE OF BACTERIAL GROWTH

Under the right laboratory conditions, a growing bacterial population doubles at regular intervals and the growth rate increases geometrically or exponentially (2^0, 2^1, 2^2, $2^3 \ldots 2^n$) where n is the number of generations. It should be noted that this explosive growth is not really representative of the growth pattern of bacteria in nature, but it illustrates the potential difficulty presented when a patient has a runaway infection, and is a useful tool in diagnosing bacterial disease.

When a medium for culturing bacteria captured from a patient in order to determine what sort of infection might be causing symptoms is inoculated with a certain number of bacteria, the culture will exemplify a growth curve similar to that illustrated below in Figure 5. Note that the growth curve is set to a logarithmic scale in order to straighten the steeply rising exponential growth curve. This works well because $\log 2^2 = 2x$ is a formula for a straight line in analytic geometry.

The bacterial growth curve displays four typical growth phases. At first there is a temporary lag as the bacteria take time to adapt to the medium environment. An exponential growth phase as described above follows as the bacteria divide at regular intervals by binary fission. The bacterial colony eventually runs out of enough nutrients or space to fuel further growth and the medium

Key Terms

Exponential growth: A growth process in which a number grows proportional to its size. Examples include viruses, animal populations, and compound interest paid on bank deposits.

Probability distribution: The expected pattern of random occurrences in nature.

becomes contaminated with metabolic waste from the bacteria. Finally, the bacteria begin to die off at a rate that is also geometric, similar to the exponential growth rate. This phenomenon is extremely useful in biomedical research because it enables investigators to culture sufficient quantities of bacteria and to investigate their genetic characteristics at particular points on the curve, particularly the stationary phase.

Potential Applications

One of the most interesting future developments in this field will likely be connected to advances in knowledge concerning the human genome that could revolutionize understanding of the pathogenesis of disease. As of 2005, knowledge of the genome has already contributed to the development of high-technology genetic screening techniques that could be just the beginning of using information about how the expression of thousands of different genes impacts the development, treatment, and prognosis of breast and other types of cancer, as well as the development of cardiovascular disease, diabetes, and other chronic diseases.

For example, researchers have identified a gene-expression profile consisting of 70 different genes that accurately predicted the prognosis for a group of breast cancer patients into poor prognosis and good prognosis groups. This profile was highly correlated with other clinical characteristics, such as age, tumor histologic grade, and estrogen receptor status. When they evaluated the predictive power of their prognostic categories in a ten-year survival analysis, they found that the probability of remaining free of distant metastases was 85.2% in the good prognosis group, but only 50.6% in the poor prognosis group. Similarly, the survival rate at ten years was 94.6% in the good prognosis group, but only 54.6% in the poor prognosis group. This result was particularly valuable because some patients that had positive lymph nodes that would have been classified as having a poor prognosis using conventional criteria were found to have good prognoses using the genetic profile.

Physicians and scientists involved in medical research and clinical trials have made enormous contributions to the understanding of the causes and the most effective treatment of disease. The most telling indicator of the impact of their work has been the steadily declining death rate throughout the world. Old challenges to human survival continue, and new ones will certainly emerge (e.g., AIDS and the diseases of obesity). The mathematical tools of medical research will continue to be humankind's arsenal in the struggle for better health.

Where to Learn More

Books

Hennekens, C.H., and J.E. Buring. *Epidemiology in Medicine.* Boston: Little, Brown & Co., 1987.

Periodicals

Coombes, R., et al. "A Randomized Trial of Exemestane after Two to Three Years of Tamoxifen Therapy in Postmenopausal Women with Primary Breast Cancer." *New England Journal of Medicine.* (March 11, 2004) 350(11): 1081–1092.

De Vincenzi,Isabelle. "A Longitudinal Study of Human Immunodeficiency Virus Transmission by Heterosexual Partners." *New England Journal of Medicine.* (August 11, 1994) 331:6: 341–346.

Fyles, A.W., et al. "Tamoxifen with or without Breast Irradiation in Women 50 Years of Age or Older with Early Breast Cancer." *New England Journal of Medicine* (2004) 351(10): 963–970.

Shapiro, S., et al. "Lead Time in Breast Cancer Detection and Implications for Periodicity of Screening." *American Journal of Epidemiology* (1974) 100: 357–366.

Van't Veer, L., et al. "Gene Expression Profiling Predicts Clinical Outcome of Breast Cancer." *Nature.* (January 2002) 415: 530–536.

Web sites

"Significant improvements in disease free survival reported in women with breast cancer." First report from The Cancer and Leukemia Group B (CALGB) 9741 (Intergroup C9741) study. December 12, 2002 (May 13, 2005). <http://www.prnewswire.co.uk/cgi/news/release?id=95527>.

"Old Breast Cancer Drug Still No. 1." WebMD, May 20, 2002. (May 13, 2005.) <http://my.webmd.com/content/article/16/2726_623.htm>.

Modeling

Overview

A model is a representation that mimics the important features of a subject. A mathematical model uses mathematical structures such as numbers, equations, and graphs to represent the relevant characteristics of the original. Mathematical models rely on a variety of mathematical techniques. They vary in size from graphs to simple equations, to complex computer programs. A variety of computer coding languages and software programs have been developed to aid in computer modeling. Mathematical models are used for an almost unlimited range of subjects including agriculture, architecture, biology, business, design, education, engineering, economics, genetics, marketing, medicine, military, planning, population genetics, psychology, and social science.

Fundamental Mathematical Concepts and Terms

There are three fundamental components of a mathematical model. The first includes the things that the model is designed to reflect or study. These are often referred to as the output, the dependent variables, or the endogenous variables. The second part is referred to as input, parameters, independent variables, or exogenous variables. It represents the features that the model is not designed to reflect or study, but which are included in or assumed by the model. The last part is the things that are omitted from the model.

Consider a marine ecologist who wants to build a model to predict the size of the population of kelp bass (a species of fish) in a certain cove during a certain year. This number is the output or the dependent variable. The ecologist would consider of all the factors that might influence the fish population. These might include the temperature of the water, the concentration of food for the kelp bass, population of kelp bass from the previous year, the number of fishermen who use the cove, and whatever else he considers important. These items are the input or the dependent variables. The things that might be excluded from the model are those things that do not influence the size of the kelp bass population. These might include the air temperature, the number of sunny days per year, the number of cars that are licensed within a 5-mile (8 km) radius of the cove, and anything else that does not have a clear, direct impact on the fish population.

Once the model is built, it can often serve a variety of purposes and the variables in the model can change depending on the model's use. Imagine that the same

model of kelp bass populations is used by an officer at the Department of Fish and Wildlife to set fishing regulations. The officer cares a lot about how many fishermen use the cove and he can set regulations controlling the number of licenses granted. For the regulator, the number of fisherman changes to the independent variable and the population of fish is a dependent variable.

Building mathematical models is somewhat similar to creating a piece of artwork. Model building requires imagination, creativity, and a deep understanding of the process or situation being modeled. Although there is no set method that will guarantee a useful, informative model, most model building requires, at the very least, the following four steps.

First, the problem must be formulated. Every model answers some question or solves a problem. Determining the nature of the problem or the fundamentals involved in the question are basic to building the model. This step can be the most difficult part of model building.

Second, the model must be outlined. This includes choosing the variables that will be included and omitted. If parameters that have no impact on the output are included in the model, it will not work well. On the other hand, if too many variables are included in the model, it will become exceedingly complex and ineffective. In addition, the dependent and independent variables must be determined and the mathematical structures that describe the relationships between the variables must be developed. Part of this step involves making assumptions. These assumptions are the definitions of the variables and the relationships between them. The choice of assumptions plays a large role in the reliability of a model's predictions.

The third step of building a model is assessing its usefulness. This step involves determining if the data from model are what it was designed to produce and if the data can be used to make the predictions the model was intended to make. If not, then the model must be reformulated. This may involve going back to the outline of the model and checking that the variables are appropriate and their relationships are structured properly. It may even require revisiting the formulation of the problem itself.

The final step of developing a model is testing it. At this point results, from the model are compared against measurements or common sense. If the predictions of the model do not agree with the results, the first step is to check for mathematical errors. If there are none, then fixing the model may require reformulations to the mathematical structures or the problem itself. If the predictions of the model are reasonable, then the range of variables for which the model is accurate should be explored. Understanding the limits of the model is part of the testing process. In some cases it may be difficult to find data to compare with predictions from the model. Data may be difficult, or even impossible, to collect. For example, measurements of the geology of Mars are quite expensive to gather, but geophysical models of Mars are still produced. Experience and knowledge of the situation can be used to help test the model.

After a model is built, it can be used to generate predictions. This should always be done carefully. Models usually only function properly within certain ranges. The assumptions of a model are also important to keep in mind when applying it.

Models must strike a balance between generality and specificity. When a model can explain a broad range of circumstances, it is general. For example, the normal distribution, or the bell curve, predicts the distribution of test scores for an average class of students. However, the distribution of test scores for a specific professor might vary from the normal distribution. The professor may write extremely hard tests or the students may have had more background in the material than in prior years. A U-shaped or linear model may better represent the distribution of test scores for a particular class. When a model more specific to a class is used, then the model loses its generality, but it better reflects reality. The trade-offs between these values must be considered when building and interpreting a model.

There are a variety of different types of mathematical models. Analytical models or deterministic models use groups of interrelated equations and the result is an exact solution. Often advanced mathematical techniques, such as differential equations and numerical methods, are required to solve analytical models. Numerical methods usually calculate how things change with time based on the value of a variable at a previous point in time. Statistical or stochastic models calculate the probability that an event will occur. Depending on the situation, statistical models may have an analytical solution, but there are situations in which other techniques such as Bayesian methods, Markov random models, cluster analysis, and Monte Carlo methods are necessary. Graphical models are extremely useful for studying the relationships between variables, especially when there are only a few variables or when several variables are held constant. Optimization is an entire field of mathematical modeling that focuses on maximizing (or minimizing) something, given a group of constraining conditions. Optimization often relies on graphical techniques. Game theory and catastrophe theory can also be used in modeling. A relatively new branch

of mathematics called chaos theory has been used to model many phenomena in nature such as the growth of trees and ferns and weather patterns. String theory has been used to model viruses.

Computers are obviously excellent tools for building and solving models. General computer coding languages have the basic functions for building mathematical models. For example, JAVA, Visual Basic and C^{++} are commonly used to build mathematical models. However, there are a number of computer programs that have been developed with the particular purpose of building mathematical models. Stella II is an object oriented modeling program. This means that variables are represented by boxes and the relationships between the variables are represented by different types of arrows. The way in which the variables are connected automatically generates the mathematical equations that build the model. MathCad, MatLab and Mathematica are based on built-in codes that automatically perform mathematical functions and can solve complex equations. These programs also include a variety of graphing capabilities. Spreadsheet programs like Microsoft Excel are useful for building models, especially ones that depend on numerical techniques. They include built-in mathematical functions that are commonly used in financial, biological, and statistical models.

Real-life Applications

Mathematical models are used for an almost unlimited range of purposes. Because they are so useful for understanding a situation or a problem, nearly any field of study or object that requires engineering has had a mathematical model built around it. Models are often a less expensive way to test different engineering ideas than using larger construction projects. They are also a safer and less expensive way to experiment with various scenarios, such as the effects of wave action on a ship or wind action on a structure. Some of these fields that commonly rely on mathematical modeling are agriculture, architecture, biology, business, design, education, engineering, economics, genetics, marketing, medicine, military, planning, population genetics, psychology, and social science. Two classic examples of mathematical modeling from the vast array of mathematical models are presented below.

ECOLOGICAL MODELING

Ecologists have relied on mathematical modeling for roughly a century, ever since ecology became an active field of research. Ecologists often deal with intricate systems in which many of the parts depend on the behavior of other parts. Often, performing experiments in nature is not feasible and may also have serious environmental consequences. Instead, ecologists build mathematical models and use them as experimental systems. Ecologists can also use measurements from nature and then build mathematical models to interpret these results.

A fundamental question in ecology concerns the size of populations, the number of individuals of a given species that live in a certain place. Ecologists observe many types of fluctuations in population size. They want to understand what makes a population small one year and large the next, or what makes a population grow quickly at times and grow slowly at other times. Population models are commonly studied mathematical models in the field of ecology.

When a population has everything that it needs to grow (food, space, lack of predators, etc.), it will grow at its fastest rate. The equation that describes this pattern of growth is $\Delta N/\Delta t = rN$. The number of organisms in the population is N, time is t, and the rate of change in the number of organisms is r. The Δ is the Greek letter delta and it indicates a change in something. The equation says that the change in the number of organisms (ΔN) during a period of time (Δt) is equal to the product of the rate of change (r) and the number of organisms that are present (N).

If the period of time that is considered is allowed to become very small and the equation is integrated, it becomes $N = N_0e^{rt}$, where N_0 is the number of organisms at an initial point in time. This is an exponential equation, which indicates that the number of organisms will increase extremely fast. Because the graph of this exponential equation shoots upward very quickly, it has a shape that is similar to the shape of the letter "J". This exponential growth is sometimes called "J-shaped" growth.

J-shaped growth provides a good model of the growth of populations that reproduce rapidly and that have few limiting resources. Think about how quickly mosquitoes seem to increase when the weather warms up in the spring. Other animals with J-shaped growth are many insects, rats, and even the human population on a global scale. The value of r varies greatly for these different species. For example, the value of r for the rice weevil (an insect) is about 40 per year, for a brown rat about 5 per year and for the human population about 0.2 per year. In addition, environmental conditions, such as temperature, will influence the exponential rate of increase of a population.

In reality, many populations grow very quickly for some time and then the resources they need to grow

become limited. When populations become large, there may be less food available to eat, less space available for each individual or predators may be attracted to the large food supply and may start to prey on the population. When this happens the population growth stops increasing so quickly. In fact, at some point, it may stop increasing at all. When this occurs, the exponential growth model, which produces a J-shaped curve, does not represent the population growth very well.

Another factor must be added to the exponential equation to better model what happens when limited resources impact a population. The mathematical model, which expresses what happens to a population limited by its resources, is $\Delta N/\Delta t = rN(1 - N/K)$. The variable K is sometimes called the carrying capacity of a population. It is the maximum size of a population in a specific environment. Notice that when the number of individuals in the population is near 0 ($N = 0$), the term $1-N/K$ is approximately equal to 1. When this is the case, the model will behave like an exponential model; the population will have rapid growth. When the number of individuals in the population is equal to the carrying capacity ($N = K$), then the term $1 - N/K$ becomes $1 - K/K$, or 0. In this case the model predicts that the changes in the size of the population will be 0. In fact, when the size of a population approaches its carrying capacity, it stops growing.

The graph of a population that has limited resources starts off looking like the letter J for small population sizes and then curves over and becomes flat for larger population sizes. It is sometimes called a sigmoid growth curve or "S-shaped" growth. The mathematical model $\Delta N/\Delta t = rN(1-N/K)$ is referred to as the logistic growth curve.

The logistic growth curve is a good approximation for the population growth of animals with simple life histories, like microorganisms grown in culture. A classic example of logistic growth is the sheep population in Tasmania. Sheep were introduced to the island in 1800 and careful records of their population were kept. The population grew very quickly at first and then reached a carrying capacity of about 1,700,000 in 1860.

Sometimes a simple sigmoidal shape is not enough to clearly represent population changes. Often populations will overshoot their carrying capacity and then oscillate around it. Sometimes, predators and prey will exhibit cyclic oscillations in population size. For example the population sizes of Arctic lynx and hare increase and decrease in a cycle that lasts roughly 9–10 years.

Ecologists have often wondered whether modeling populations using just a few parameters (such as the rate of growth of the population, the carrying capacity) accurately portrays the complexity of population dynamics. In 1994, a group of researchers at Warwick University used a relatively new type of mathematics called chaos theory to investigate this question.

A mathematical simulation model of the population dynamics between foxes, rabbits and grass was developed. The computer screen was divided into a grid and each square was assigned a color corresponding to a fox, a rabbit, grass, and bare rock. Rules were developed and applied to the grid. For example, if a rabbit was next to grass, it moved to the position of the grass and ate it. If a fox was next to a rabbit, it moved to the position of the rabbit and ate it. Grass spread to an adjacent square of bare rock with a certain probability. A fox died if it did not eat in six moves, and so on.

The computer simulation was played out for several thousand moves and the researchers observed what happened to the artificial populations of fox, rabbits, and grass. They found that nearly all the variability in the system could be accounted for using just four variables, even though the computer simulation model contained much greater complexity. This implies that the simple exponential and logistic modeling that ecologists have been working with for decades may, in fact, be a very adequate representation of reality.

MILITARY MODELING

The military uses many forms of mathematical modeling to improve its ability to wage war. Many of these models involve understanding new technologies as they are applied to warfare. For example, the army is interested in the behavior of new materials when they are subjected to extreme loads. This includes modeling the conditions under which armor would fail and the mechanics of penetration of ammunition into armor. Building models of next generation vehicles, aircraft and parachutes and understanding their properties is also of extreme importance to the army.

The military places considerable emphasis on developing optimization models to better control everything from how much energy a battalion in the field requires to how to get medical help to a wounded soldier more effectively. Special probabilistic models are being developed to try to detect mine fields in the debris of war. These models incorporate novel mathematical techniques such as Bayesian methods, Markov random models, cluster analysis, and Monte Carlo simulations. Simulation models are used to develop new methods for fighting wars. These types of models make predictions about the outcome of war since it has changed from one of battlefield combat to one that incorporates new technologies like smart weapon systems.

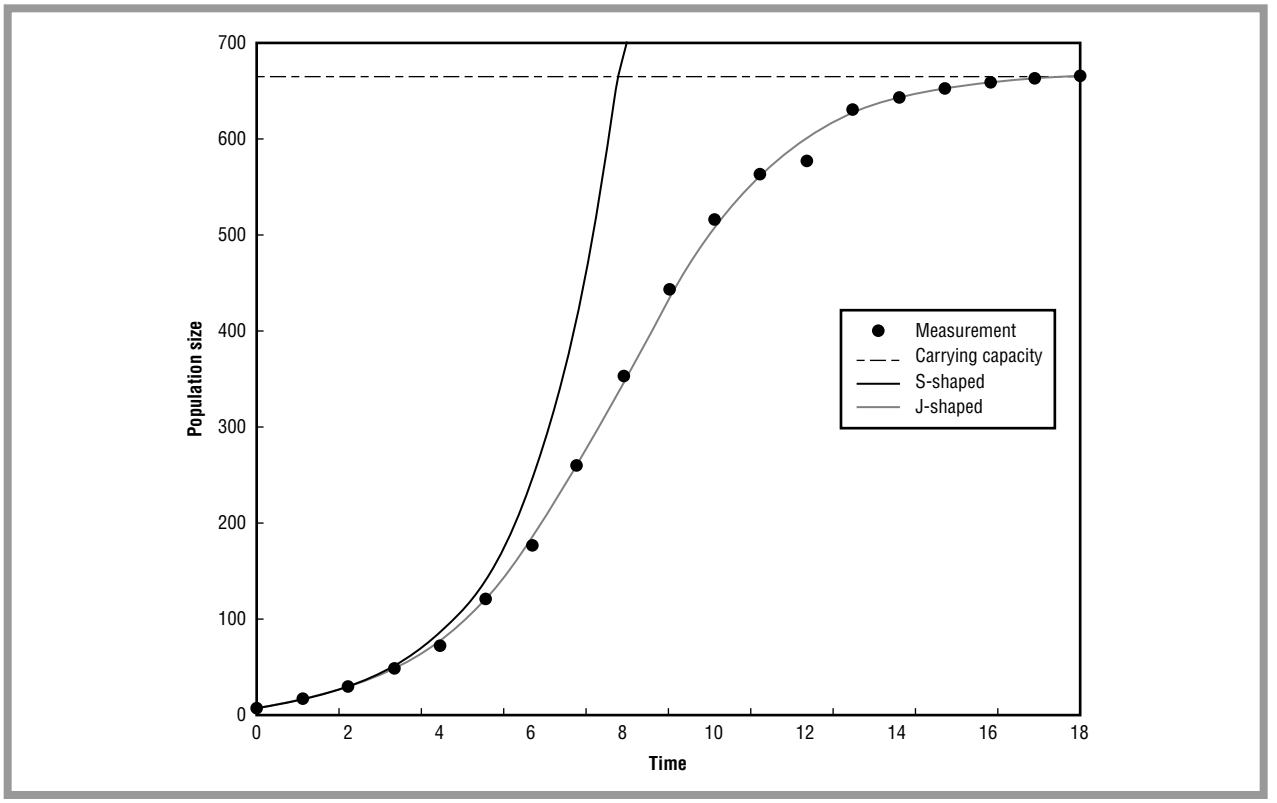


Figure 1: Examples of population growth models. The dots are measurements of the size of a population of yeast grown in a culture. The dark line is an exponential growth curve showing J-shaped growth. The lighter line is a sigmoidal or logistic growth curve showing S-shaped growth. The dashed line shows the carrying capacity of the population.

Game theory was developed in the first half of the twentieth century and applied to many economic situations. This type of modeling attempts to use mathematics to quantify the types of decisions a person will make when confronted with a dilemma. Game theory is of great importance to the military as a means for understanding the strategy of warfare. A classic example of game theory is illustrated by the military interaction between General Bradley of the United States Army and General von Kluge of the German Army in August 1944, soon after the invasion of Normandy.

The U.S. First Army had advanced into France and was confronting the German Ninth Army, which outnumbered the U.S. Army. The British protected the U.S. First Army to the North. The U.S. Third Army was in reserve just south of the First Army.

General von Kluge had two options; he could either attack or retreat. General Bradley had three options concerning his orders to the reserves. He could order them to the west to reinforce the First Army; he could order them to the east to try to encircle the German Army; or he could order them to stay in reserve for one day and then order them to reinforce the First Army or strike eastward against the Germans.

In terms of game theory, six outcomes result from the decisions of the two generals and a payoff matrix is constructed which ranks each of the outcomes. The best outcome for Bradley would be for the First Army's position to hold and to encircle the German troops. This ranks 6, or the highest in the matrix and it would occur if von Kluge attacks and the First Army and Bradley holds the Third Army in reserve one day to see if the First Army needed reinforcement and if not he could then order them to the east to encircle the German troops. The worst outcome for Bradley is a 1 and it would occur if von Kluge orders an attack and at the same time Bradley ordered the reserve troops eastward. In this case, the Germans could possibly break through the First Army's position and there would be no troops available for reinforcement.

Game theory suggests that the best decision for both generals is one that makes the most of their worst possible

Bradley's options	Von Kluge's Options: Attack	Von Kluge's Options: Retreat
Order reserves west	Outcome: Reserves reinforce First Army. Hold position Rank: 3	Outcome: Reserves reinforce First Army. Not available to harass German retreat Rank: 2
Order reserves east	Outcome: Reserves not available to reinforce First Army. Germans break through position. Rank: 1	Outcome: Reserves not available to reinforce First Army, but can harass German retreat. Rank: 5
Hold reserves one day then move west to reinforce or move east to encircle	Outcome: Reserves available to reinforce First Army if neede. If not, reserves can move to west possibly encircle German Army. Rank: 6	Outcome: Reserves available to put heavy pressure on German retreat. Rank: 4

Military positions:

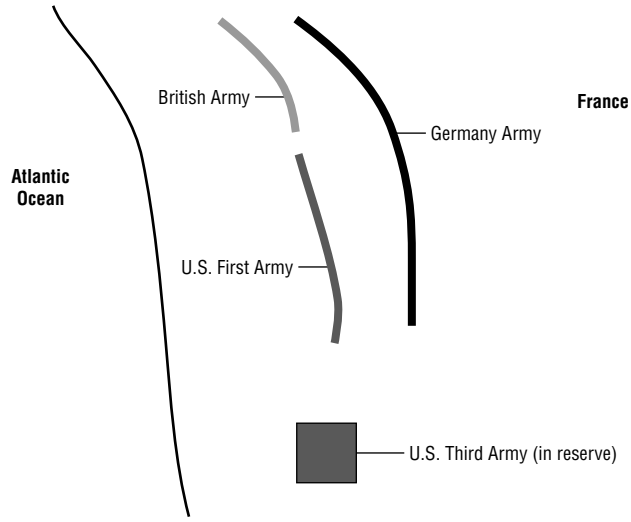


Figure 2: Payoff matrix for the various scenarios in the battle between the U.S. Army and the German Army in 1944. If possible add graphic of military positions as well. Caption should read: Military positions of the U.S. and German Armies during the battle. The U.S. and British forces held positions to the west of the German Army. The U.S. Third Army was in reserve to the south of the U.S. First Army.

outcome. Given the six scenarios, this results in von Kluge deciding to withdraw and Bradley deciding to hold the Third Army in reserve for one day, a 4 in the matrix. The expected outcome of this scenario is that the Third Army would be one day late in moving to the east and could only put moderate pressure on the retreating German Army. On the other hand, they would not be committed to the wrong action. From the German point of view, the Army does not risk being encircled and cut off by the Allies, and it avoids excessive harassment during its retreat.

Interestingly, the two generals decided to follow the action suggested by game theory. However, after van Kluge decided to withdraw, Hitler ordered him to attack. The U.S. First Army held their position on the first day of

Key Terms

Dependent variable: What is being modeled; the output.

Exponential growth: A growth process in which a number grows proportional to its size. Examples include viruses, animal populations, and compound interest paid on bank deposits.

Independent variable: Data used to develop a model, the input.

Input: What is used to develop a model, the independent variables.

Model: A system of theoretical ideas, information, and inferences presented as a mathematical description of an entity or characteristic.

Output: What is being modeled; the dependent variables.

the battle and Bradley ordered the Third Army to the east to encircle the Germans. Hitler unwittingly generated the best possible outcome for Bradley, the 6th or highest rank in the matrix.

Where to Learn More

Books

Beltrami, Edward. *Mathematical Models in the Social and Biological Sciences*. Boston: Jones and Bartlett Publishers, 1993.

Bender, Edward A. *An Introduction to Mathematical Modeling*. Mineola NY: Dover Publications, 2000.

Burghes, D.N., and A.D. Wood. *Mathematical Models in the Social, Management and Life Sciences*. Chichester: Ellis Horwood Limited, 1980.

Harte, John. *Consider a Spherical Cow*. Sausalito CA: University Science Books, 1988.

Odum, Eugene P. *Fundamentals of Ecology*. Philadelphia: Saunders College Publishing, 1971.

Skiena, Steven.*Calculated Bets: Computers, Gambling, and Mathematical Modeling to Win*. Cambridge: Cambridge University Press, 2001.

Stewart, Ian. *Nature's Numbers*. New York: BasicBooks, 1995.

Web sites

Carlton College. "Mathematical Models." Starting Point: Teaching Entry Level Geoscience. January 15, 2004. <http://serc.carleton.edu/introgeo/models/mathematical/> (April 18, 2005).

Department of Mathematical SciencesUnited States Military Academy. "Military Mathematical Modeling (M3)" May 1, 1998. <http://www.dean.usma.edu/departments/math/pubs/mmm99/default.htm> (April 18, 2005).

Multiplication

Overview

Multiplication is a method of easily adding various quantities of identical numbers without performing each addition equation individually.

Fundamental Mathematical Concepts and Terms

In a multiplication equation, the two values being multiplied are called coefficients or factors, while the result of a multiplication equation is labeled the product. Several forms of notation can be used to designate a multiplication operation. The most common symbol for multiplication in arithmetic is $\times$. In algebra and other forms of mathematics where letters substitute for unknown quantities, the $\times$ is often omitted, so that the expression $3x + 7y$ is understood to mean $3 \times x + 7 \times y$. In other cases, parentheses can be used to express multiplication, as in $5(2)$, which is mathematically identical to 5×2, or 10.

For both subtraction and division, the order of the values being operated on has a significant impact on the final answer; in multiplication, the order has no effect on the result. The commutative property of multiplication states that $x \times y$ gives the same result as $y \times x$ for any values of x and y, making the order of the factors irrelevant to the product. Another property of multiplication is that any value multiplied times 0 produces a product of 0, while any number multiplied times 1 gives the starting number. The signs of the factors also affect the product; multiplying two numbers with the same sign (either two positives or two negatives) will produce a positive result, while multiplying numbers with differing signs will produce a negative value.

A Brief History of Discovery and Development

As an extension of the basic process of addition, multiplication's origins are lost in ancient history, and early merchants probably learned to perform basic multiplication operations long before the system was formalized. The first formal multiplication tables were developed and used by the Babylonians around 1800 B.C. One of these earliest tables was created to process simple calculations of the area of a square farm field, using the length and width as data and allowing a user to look up the area in the table body. These early tables function identically to

Girl executing simple multiplication problems. Lambert/Getty Images.

today's multiplication tables, meaning that the tables which modern elementary school students labor to memorize have actually been in use for close to forty centuries.

Moving past the basic single digit equations of the elementary school multiplication table, long multiplication can become a time-consuming, complex process, and many different techniques for performing long multiplication have been developed and used. In the thirteenth century, educated Europeans used a multiplication technique known as lattice multiplication. This somewhat complicated method involved drawing a figure resembling a garden lattice, then writing the two factors above and to the right of the figure. Following a step by step process of multiplying, adding, and summing values, this method allowed one to reliably multiply large numbers.

An earlier, more primitive method of long multiplication was devised by the early Egyptians, and is described in a document dating to 1700 B.C. The Egyptian system seems rather unusual, due largely to the Egyptian perspective on numbers. Whereas modern mathematics views numbers as independent, discrete entities with an inherent value, ancient Egyptians thought of numbers only in terms of collections of concrete objects. In other words, to an ancient Egyptian, the number nine would have no inherent meaning, but would always refer to a specific collection of objects, such as nine swords or nine cats.

For this reason, Egyptian math generally did not attempt to deal with extremely large quantities, as these calculations offered little practical value. Instead, the Egyptians devised a method of multiplication which could be accomplished by a complex series of manipulations using nothing more than simple addition. Due to its complexity and limited utility, this method does not appear to have gained favor outside Egypt. As an interesting side note, elements of the Egyptian method actually involve binary mathematics, the system which forms the basis of modern computer logic systems.

A similar, binary-based system was developed and used in Russia. This so-called peasant method of multiplication involved repeatedly doubling and halving the two values to be multiplied until an answer was produced. While tedious to apply, this method involved little more than removing the right-most value at each step until the result was produced. Like the previously discussed methods, this technique seems remarkably slow in modern terms; however, in a context in which one might only need to perform a single multiplication problem each week or each month, such techniques would have been useful.

Given the complexity of performing long multiplication manually, numerous inventors attempted to create mechanical multiplying machines. Far more difficult than creating a simple adding machine, this task was first successfully completed by Gottfried Wilhelm Von Leibniz (1646–1716), a German philosopher and mathematician who also invented differential calculus. This device, which Von Leibniz called the Stepped Reckoner, used a series of mechanical cranks, drums, and gears to evaluate multiplication equations, as well as division, addition, and subtraction problems. Only two of these machines were ever built; both survive and are housed in German museums. Von Leibniz apparently shared the somewhat common dislike of calculating by hand; he is quoted as saying that the process of performing hand calculations squanders time and reduces men to the level of slaves. Unfortunately, his bulky, complex mechanical calculator never came into widespread use.

Additional attempts were made to construct multiplying machines, and various mechanical and electromechanical versions were created during the ensuing centuries. However the first practical hand-held tools for performing multiplication did not appear until the 1970s, with the introduction of microprocessors and handheld calculators by firms such as Hewlett Packard and Texas Instruments. Today, using these inexpensive

tools or spreadsheet software, long multiplication is no more difficult or time-consuming to perform than simple addition.

Real-life Applications

EXPONENTS AND GROWTH RATES

Growth rates describe the application of simple multiplication many times to a starting value. In cases where the growth rate is constant over time, a special notation is used to define the projected value; this notation is called an exponent, and its value conveys how many times the starting value is to be multiplied by itself. For example, the expression 3×3 can also be written 3^2, which is read "Three to the second power," or simply "Three squared." As the sequence progresses, the values become more cumbersome to work with, and exponents greatly simplify the process. For instance, the expression $3 \times 3 \times 3 \times 3 \times 3 \times 3 \times 3 \times 3 \times 3 \times 3$ can be easily written as 3^{10}, and when evaluated produces a value of 59,049.

INVESTMENT CALCULATIONS

One common application of exponents deals with growth rates. For example, assume that an investment of \$100 will earn 7% over the course of a year, yielding a total of \$107 at year-end. This process can be continued indefinitely; at the end of two years, this \$107 will have earned another \$7.49, making the total after two years \$114.49.

Using exponents, we can easily determine how much the original \$100 will have earned after any specific number of years; in this example, we will find the total value after nine years. First, we note that the growth rate is 7%, meaning that the starting value must be multiplied by 1.07 in order to find the new value after one year. In order to find the multiplier, or value we would apply to our starting number to find the final total, we simply multiply 1.07 times itself until we account for all nine years of the calculation. Expressed in long terms, this equation would be $1.07 \times 1.07 \times 1.07 \times 1.07 \times 1.07 \times 1.07 \times 1.07 \times 1.07 \times 1.07 = 1.84$. Expressing this value exponentially we write the expression as 1.07^9. We can now multiply our original investment value by our calculated multiplier to find the final value of the investment: $\$100 \times 1.84 = \$184$. Further, if we wish to recalculate the size of the investment over a shorter or longer period of time, we simply change the exponent to reflect the new time period.

Two unusual situations occur when using exponents. First, by convention, the value of any number raised to the power 0 is 1; so $4^0 = 1$, $26^0 = 1$, and $995^0 = 1$. While mathematicians offer lengthy explanations of why this is so, a more intuitive explanation is simply that moving from one exponent to the next lower one requires a division by the base value; for example, to move from 3^4 to 3^3, we divide by 3, or in expanded terms, we divide 81 by 3 to get to 27. If we follow this sequence to its natural progression, we will eventually reach 3^1, and if we divide this value (3) by 3, we find a result of 1. Since this sequence will end with 1 for any base value, then any value raised to the power 0 will equal 1.

A second curiosity of exponents occurs in the case of negative values, either in the exponent or in the base value. In some situations, base values are raised to a negative power, as in the expression 5^{-3}. By convention, this

How Much Wood Could a Woodchuck Chuck, if a Woodchuck Could Chuck Wood?

This nursery rhyme tongue-twister has puzzled children for years, and has in fact inspired numerous online discussions regarding the specific details of the riddle and how to solve it. Using a simple formula, we can take the amount the rodent chucks per hour, multiply it times the number of working hours each day, then multiply again by 365 to get a total per year. This, multiplied by the animal's lifespan in years would give us a total amount chucked, which one online estimate places at somewhere around 26 tons.

Like all such estimations, in which a single event is multiplied repeatedly to predict performance over a long period of time, this estimate is fraught with assumptions, any of which can cause the final estimate to be either too high or too low. For example, even a small error in estimating how much can be chucked per hour could throw the final total off by a ton or more. Another major source of error is found in the variability of the woodchuck's work; unlike mechanical wood chuckers, woodchucks work faster some days than others. Also unlike machines, rodents frequently spend the winter hibernating, significantly reducing the actual volume of wood chucked. To sum up, the question of how much wood can be chucked remains difficult to answer, given the number of assumptions required; the most generally correct answer may simply be "Quite a lot."

expression is evaluated as the inverse of this expression with the exponent sign made positive, or $1/5^3 = 1/125$. A related complication arises when the base value is itself negative, as in the case of $(-5)^3$. Multiplying negative and positive values is accomplished according to a simple set of rules: if the signs are the same, the final value is positive, otherwise the final value is negative. So 4×4 and -4×-4 produce the same result, a value of 16. However 4×-4 produces a value of –16. In the case of a negative base being raised to a specific power, a related set of rules apply: if the exponent is even, the final value is positive, otherwise it is negative. Following this rule, $(-5)^3 = -125$, while $(-5)^2 = 25$.

CALCULATING EXPONENTIAL GROWTH RATES

One ancient myth is based on the concept of an exponential growth rate. The legend of Hercules describes a series of twelve great works which the hero was required to perform; one of these assignments was to slay the Hydra, a horrible beast with nine heads. While Hercules was unimaginably strong, he quickly found that traditional tactics would not work against the Hydra; each time Hercules cut off one of the Hydra's heads, two new heads grew in its place, meaning that as soon as he turned from dispatching one head, he quickly found himself being attacked by even more heads than before. Hercules eventually triumphed by discovering how to cauterize the stumps of the severed heads, preventing them from regenerating. While this story is ancient, it illustrates a simple principle which frequently holds true: in stopping an exponentially growing system, the best solution is typically to interrupt the growth cycle, rather than trying to keep up with it in this case was to prevent or interrupt the growth in the first place, rather than trying to keep up with it as it occurs.

While some animals are able to regenerate severed body parts, no real-life animal is able to do so as quickly as the mythical Hydra. However, some animal populations do multiply at an alarming rate, and in the right circumstances can rapidly reach plague proportions. Mice, for example, can produce offspring about every three weeks, and each litter can include up to eighteen young. To simplify this equation, we can assume one litter per month, and 16 young per litter. We also assume, for simplicity, that the mice only live to be 1 month old, so only their offspring live on into the next month. Beginning with a single pair of healthy mice on New Year's Day, by the end of January, we will have eight pair. Thus, over the course of the first month, the mouse population will have grown by a factor of eight.

While this first month's performance is impressive, the process becomes even more startling as the months pass. At the end of February, the eight pair from month one will have each given birth to another sixteen young (eight pair), making the new population $8 \times 8 = 64$ pair. This number will continue to increase by a factor of eight each month, meaning that by the end of May, more than 3,000 pair of mice will exist. By the end of December, the total mouse population will be almost 70 billion, or about 10 times the human population of Earth.

Obviously, mice have lived on Earth for eons without ever taking over, so this conclusion raises some question about the validity of the math involved, as well as pointing out some potential problems with the methodology used. First, the calculation assumes that mice can begin breeding immediately after birth, which is incorrect. Also, it assumes that all the mice in each generation survive to reproduce, when in fact many mice do not. Additionally, it assumes that adequate food exists for all the mice to continue eating, which would also be a near-impossibility. Finally, it assumes that the mouse's natural predators, including humans, would sit idly by and watch this takeover occur. Since these limitations all impact the growth rate of mouse populations in real life, a mouse population explosion of the size described here is unlikely to occur. Nevertheless, the high multiplication rate of mice and other rodents helps explain why they are so difficult to eradicate, and can so quickly infest large buildings.

While the final result of the mouse calculation is somewhat unrealistic, similar population explosions have actually occurred. A small number of domestic rabbits were released in Australia during the 1800s; with adequate food and few natural predators, they quickly multiplied and began destroying the natural vegetation. During the 1950s, government officials began releasing the Myxoma virus, which killed 99% of animals exposed to it. However, resistant animals quickly replenished the population, and by the mid-1990s, parts of the Australian rangeland were inhabited by more than 3,000 rabbits per square kilometer. Rabbit control remains an issue in Australia today; the country boasts the world's longest rabbit fence, which extends more than 1,000 kilometers. As of 1991, the estimated rabbit population of Australia was approximately 300 million, or about fifteen times the human population of the continent.

SPORTS MULTIPLICATION

CALCULATING A BASEBALL ERA

Comparing the performance of baseball pitchers can be difficult. In a typical major league game three, four, or more pitchers all work for the same goal, but only one is

awarded a win or loss. To help compare pitching performance on more even basis, baseball analysts frequently discuss a pitcher's earned run average, or ERA. The ERA is used to evaluate what might happen if pitchers could pitch entire games, providing a basis for comparison among multiple players.

Calculating a pitcher's ERA is fairly simple, and involves just a few values. The process begins with the number of earned runs scored on the pitcher during his time in the game. This value is then multiplied by nine (the assumed number of innings in a full game), and that total is divided by the number of innings actually pitched. For example, if a pitcher plays three innings and allows two runs, his ERA would be calculated as $2 \times 9/3 = 6$. Like most projections, this one is subject to numerous other factors, but suggests that if this pitcher could maintain his performance at this level, he would allow six runs in a typical full game.

The ERA calculation becomes more complex when a pitcher is removed from a game during an inning. In such cases, the number of innings pitched will be measured in thirds, with each out equaling one third of an inning. If the pitcher who allows two runs is removed after one out has been made in the fourth inning, he would have pitched 3 1/3 innings. Historically, major league ERAs have risen and fallen as the rules of the game have changed. Today, a typical professional pitcher will have an ERA around 4.50, while league leaders often post single-season ERAs of 2.00 or less. One of a coach's more difficult challenges is recognizing when a pitcher has reached the end of his effectiveness and should be removed from a game. Fatigue typically leads to poorer performance and a rapidly rising ERA.

RATE OF PAY

An old joke says that preachers hold the most lucrative jobs, since they are paid for a week's labor but only work one day of each week. Using this arguably flawed logic, professional rodeo cowboys might be considered some of the highest paid athletes today, since they spend so little time actually "working." A bull rider's working weekend typically consists of a two day competition. Each competitor rides one bull the first night, and a second the following night. If he is able to stay on each bull for the full eight seconds, and scores enough style points for his riding ability, he then qualifies for a third ride in the final round of competition.

Because each ride lasts only eight seconds, a bull rider's complete work time for each event is only 24 seconds, not counting time spent settling into the saddle and the inevitable sprint to escape after the ride ends. Multiplying this 24 seconds of work times the 31 events in an entire professional season produces a total working time each year of about 13 minutes. Because a top professional rider earns over $250,000 per season, this rider's income works out to an amazing $19,230 per minute, or $1,153,846 per hour. Unfortunately, this average does not include the enormous amounts of time spent practicing, traveling, and healing from injuries, and in many cases, professional bull riders win only a few thousand dollars per season. But even for the wages paid to top riders, few people are willing to strap themselves atop an angry animal that weighs more than some small cars.

MEASUREMENT SYSTEMS

Some sports have their own unique measurement systems. Horse racing is a sport in which races are frequently measured in furlongs; since a furlong is approximately 66 feet, a 50 furlong race would be 3,300 feet long, or around .6 miles. Furlongs can be converted to feet by multiplying by 66, or converted to miles by dividing by 80. Horses themselves are frequently measured in an arcane measurement unit, the hand. A hand equals approximately four inches, and hands can be converted to feet by multiplying the number of hands by 3, or to inches by multiplying the number of hands by .25. Like many other traditional units of measurement, the hand is a standardized version of an ancient method of measurement, in which the width of four fingers serves as a standard measurement tool.

ELECTRONIC TIMING

Electronic timing has made many sports more exciting to watch, with Olympic medallists often separated from also-rans by mere thousandths of a second. In some events, split times are calculated, such as a time at the halfway mark of a downhill ski race. Along with providing an assessment of how well a skier is performing on the top half of the course, these measurements can also be used to predict the final time by simply doubling the mid-point time to predict the final. While this method is not foolproof, it is close enough to give fans an idea of whether a skier will be chasing a world record or simply trying to reach the bottom of the hill without falling down.

MULTIPLICATION IN INTERNATIONAL TRAVEL

Despite enormous growth in international trade, the United States still uses the imperial measurement system, rather than the more common and simpler metric system.

Because of this disparity, conversions between the two systems are sometimes necessary. While the 2-liter soft drink is one of the few common uses of the metric system in America today, a short trip to Canada would reveal countless situations in which converting between the two systems would be necessary.

While packing for the trip, an important consideration would be the weather forecast for Canada, which would normally be given in degrees Celsius. The conversion from Celsius to the Fahrenheit system used in the U.S. requires multiplication and division, using this formula: $F = 9/5 \times C + 32$. To get a ballpark figure (a rough estimate), simply double the Celsius reading and add 30. Obviously, this difference in measurement systems means that a frigid sounding temperature of 40 degrees Celsius is in fact quite hot, equal to 104 degrees Fahrenheit. Converting Fahrenheit to Celsius is equally simple: just reverse the process, subtracting 32 and multiplying by 5/9. No conversion is necessary at –40, because this is the point at which both scales read the same value.

Driving in Canada would also require mathematical conversions; while Canadians drive on the right-hand side of the highway, they measure speed in kilometers per hour (km/h), rather than the U.S. traditional miles per hour (mph) system. Because one mile equals 1.6 kilometers, the kilometer values for a given speed are larger than the mile values; the typical highway speed of 55 mph in the U.S. is approximately equal to 88 km/h in Canada, and mph can be converted to km/h using a multiplication factor of 1.6.

Gasoline in Canada is often more expensive than in the United States; however prices there are not posted in gallons, but in liters, meaning the posted price may appear exceptionally low. One gallon equals 3.8 liters, and gallons are converted to liters by multiplying by this value. Soft drinks are often sold in 2-liter bottles in the U.S., making this one of the few metric quantities familiar to Americans. Also, smaller volumes of liquid are measured not in ounces, quarts, or pints, but in deciliters and milliliters.

One of the greatest advantages of the metric system is its simplicity, with unit conversions requiring only a shift of the decimal point. For example, under the U.S. system, converting miles to yards requires one to multiply by 1,760, and converting to feet requires multiplication by 5,280. Liquids are even more confusing, with gallons to quarts using a factor of 4, and quarts to ounces using 32. Weights are similarly inconsistent, with pounds equaling 16 ounces. Using the metric system, each conversion is based on a factor of ten: multiplying by ten, one hundred, or one thousand allows conversions among kilometers, meters, and millimeters for distance, liters, deciliters, and milliliters for volume, and kilograms, decigrams, and milligrams for weight.

OTHER USES OF MULTIPLICATION

Multiplication is frequently used to find the area of a space; as previously discussed, one of the oldest known multiplication tables was apparently created to calculate the total area of pieces of farm property based on only the side dimensions. The area of a square or rectangle is found by multiplying the length times the width; for a field 40 feet long and 20 feet wide, the total area would be $40 \times 20 = 800$ square feet. Other shapes have their own formulae; a triangle's area is calculated by multiplying the length of the base by the height, then multiplying this total by 0.5; a triangle with a 40 foot base and a 20 foot height would be half the size of the previously described rectangle, and its area would be $40 \times 20 \times 0.5 = 400$ square feet.

Formulas also exist for determining the area of more complex shapes. While simple multiplication will suffice for squares, rectangles, and triangles, additional information is needed to find the area of a circle. One of the best-known and most widely used mathematical constants is the value pi, which is approximately 3.14. Pi was first calculated by the ancient Babylonians, who placed its value at 3.125; in 2002, researchers calculated the value of pi to the 1.2 trillionth decimal place.

Pi's value lies in its use in calculating both the circumference and the area of a circle. The circumference, or distance around the perimeter, of a circle, is found by multiplying pi times the diameter; for a circle with diameter of 10 inches, the circumference would be 3.14×10, or 31.4 inches. The area of this same circle can be found by multiplying pi times the radius squared; for a circle with diameter of 10 and radius of 5, the formula would be $3.14 \times 5 \times 5$, giving an area of 78.5 square inches.

Other techniques can be used to calculate the area of irregular shapes. One approach involves breaking an irregular shape into a series of smaller shapes such as rectangles and triangles, finding the area of each smaller shape, and adding these values together to produce a total; this method is frequently used when calculating the number of shingles needed to cover an irregularly shaped roof.

A branch of mathematics called calculus can be used to calculate the area under a curve using only the formula which describes the curve itself. This technique is fundamentally similar to the previously described method, in that it mathematically slices the space under the curve into extremely thin sections, then finds the area of each

and sums the results. Calculus has numerous applications in fields such as engineering and physics.

CALCULATING MILES PER GALLON

As the price of gasoline rises and occasionally falls, one common question deals with how to reduce the cost of fuel. The initial part of this question involves determining how much gas a car uses in the first place. Some cars now have mileage computers which calculate this automatically, but for most drivers, dividing the number of miles driven (a figure taken from the trip odometer) by the number of gallons added (a figure on the fuel pump) will provide a simple measure of miles per gallon. Using this figure along with the capacity of the fuel tank allows a calculation of a vehicle's range, or how far it can travel before refueling.

In general, larger vehicles will travel fewer miles per gallon of gas, making them more expensive to operate. However, these vehicles also typically have larger fuel tanks, making their range on a single tank equal to that of a smaller car. For example, a 2003 Hummer H2 has a 30-gallon fuel tank and gets around 12 miles per gallon, giving it a theoretical range of 360 miles on a full tank. In comparison, the fuel-sipping 2004 Toyota Prius hybrid sedan has only a 12 gallon tank. However, when combined with the car's mileage rating of more than 50 miles per gallon, this vehicle can travel around 600 miles per tank, and could conceivably travel more than 1,500 miles on the Hummer's oversized 30-gallon fuel load. In general, most cars are built to allow a 300–500-mile driving range between fill-ups, however the price of the fill-up varies widely depending on the car's efficiency and tank size.

SAVINGS

Small amounts of money can often add up quickly. Consider a convenience store, and a student who stops there each morning to purchase a soft drink. These drinks sell for $1.00, but by reusing his cup from previous days, the student could save 32 cents per day, since the refill price is only 68 cents. While this amount of money seems trivial when viewed alone, consider the implications over time.

Over the course of just one week, this small savings rapidly adds up; multiplying the savings times five days gives a total savings of $1.60, or enough to buy two more refills. Multiplying this weekly savings times four gives us a monthly savings of around $6.40, and multiplying the weekly savings by 52 yields a total annual savings of $83.20, enough to pay for a tank or two of gas or perhaps a nice evening out. Perhaps more amazing is the result when a consumer decides to save small amounts wherever possible; saving this same tiny amount on ten items each day would yield annual savings of $832.00, a significant amount of savings for doing little more than paying attention to how the money is being spent.

Potential Applications

One increasingly popular marketing technique illustrates the use of exponential growth for practical use. Traditional marketing practices work largely by addition: as more advertisements are run, the number of potential customers grows and a percentage of those potential customers eventually buy the product or service. As advertising markets have become more fragmented and audiences have grown harder to reach, one emerging technique is called viral marketing.

SPAM AND EMAIL COMMUNICATIONS

Viral marketing refers to a marketing technique in which information is passed from the advertiser to one generation of customers who then pass it to succeeding generations in rapidly expanding waves. In the same way that the rabbit population in Australia expanded by several times as each generation was born, viral marketing depends on people's tendency to pass messages they find amusing or thought-provoking to a long list of friends.

The growth of e-mail in particular has helped spur the rise of viral marketing, since forwarding a funny email is as simple as clicking an icon. In the same way that viruses rapidly multiply, viral e-mail messages can expand so rapidly that they clog company e-mail servers. Some companies have begun taking advantage of this phenomenon by intentionally producing and releasing viral marketing messages, such as humorous parodies of television commercials. Viral marketing can be an exceptionally inexpensive technique, as the material is distributed at no cost to the originating firm.

Where to Learn More

Web sites

Brain Bank. "A History of Measurement and Metrics." <http://www.cftech.com/BrainBank/OTHERREFERENCE/WEIGHTSandMEASURES/MetricHistory.html> (April 9, 2005).

A Brief History of Mechanical Calculators. "Leibniz Stepped Drum." <http://www.xnumber.com/xnumber/mechanical1.htm> (April 5, 2005).

Centre for Experimental and Constructive Mathematics. "Table of Computation of Pi from 2000 B.C. to Now." <http://

Key Terms

Exponential growth: A growth process in which a number grows proportional to its size. Examples include viruses, animal populations, and compound interest paid on bank deposits.

Integral calculus: A branch of mathematics used for purposes such as calculating such values as volumes displaced, distances traveled, or areas under a curve.

www.cecm.sfu.ca/projects/ISC/Pihistory.html> (April 5, 2005).

Ferel feast!. "History of Rabbits in Australia." <http://library.thinkquest.org/03oct/00128/en/rabbits/history.htm> (April 7, 2005).

History of Electronics. "Calculators." <http://www.etedeschi.ndirect.co.uk/museum/concise.history.htm> (April 6, 2005).

Homerun Web. "How to Calculate Earned Run Average (ERA)." <http://www.homerunweb.com/era.html> (April 8, 2005).

Information Technology at St. Lawrence University. "Early Dynastic Mathematics." <http://it.stlawu.edu/~dmelvill/mesomath/3Mill/ED.html> (April 6, 2005).

The Math Forum at Drexel. "Egyptian Method of Multiplication." <http://mathforum.org/library/drmath/view/57542.html> (April 5, 2005).

The Math Forum at Drexel. "Lattice Multiplication." <http://mathforum.org/library/drmath/view/52468.html> (April 4, 2005).

Music and Mathematics

Overview

Mathematics and music are basic elements of cultures and civilizations. They are fundamental. Along with moving, speaking, and reading, basic mathematics is one of the key early developmental skills parents try to instill in their children. Even before children are born, their parents may play music for them. Music is said to help babies' brains develop, and music made specifically for unborn or newborn babies can be found ranging from classical to reggae. Mathematics and music are often combined in books, toys, or songs, as musical counting can be easier for children to remember and is more fun.

Fundamental Mathematical Concepts and Terms

Mathematics is the study of mathematics. Music is experienced and created as music. These statements may seem obvious and trivial at first, but consider the study of physics without mathematics, or the study of economics without statistics, or the creation of poetry without language. Most subjects draw on tools from other disciplines, but mathematics and music can be studied as pure forms. In music, the form and the medium are the same. In mathematics, the methods and the subject are the same.

What is a number? What is a minor key? Both mathematics and music have invented special symbols to represent their seemingly abstract concepts. Everyone thinks they know what a number is, but defining the concept of a number is very difficult. One may know what a minor key is, but defining the concept rigorously is again difficult.

However, these are casual similarities between mathematics and music. There are more fundamental and formal ways in which the two disciplines interact, and the realization of this dates back to at least the Ancient Greeks.

A Brief History of Discovery and Development

PYTHAGORAS AND STRINGS

Pythagoras (fifth to sixth century B.C.) is chiefly remembered for discovering the method of calculating the length of one side of a right-angled triangle when the lengths of the other sides are known. He was a Greek philosopher who came to believe that mathematics was the most important discipline to study, and that nature was, at its deepest level, mathematical. Very little is known

about Pythagoras, and what writings there are have come from his later followers and contain many inventions and exaggerations, often treating Pythagoras as a god-like being with divine powers. However, even if nothing his supporters tell is true, the legend and legacy of Pythagoras had a profound effect on both mathematics and music in the Western tradition.

Supposedly, one day Pythagoras was walking by a blacksmith's shop and was distracted by the sounds of the hammers falling on the anvil. Several hammers were being used, producing distinct notes, the notes separated in regular musical intervals that Pythagoras found pleasing to hear. The hammers were of different weights, and Pythagoras wondered if the ratio of weights might be related to the notes they played. Then Pythagoras is said to have done something unusual for a Greek philosopher; he experimented to see if his observations had a physical basis, by playing with a string.

The string was a monochord, which is a taut string stretched between two supports on the top of a hollow sounding box, much like a single string guitar. Plucking the string makes it vibrate, which produces a note dependent on the length and tension of the string. A stopper, or bridge, can move up and down the string, changing the length of the string that vibrates, and thereby changing the note played. The shorter the string, the higher the note.

Only certain combinations of notes seem to go together to the ear. Pythagoras had found the music of the hammers pleasing. The notes they played when striking the anvil seemed to complement each other. The monochord, like many stringed instruments, can play notes that seem 'sweet' and those that seem 'sour' to the ear, as well as those that seem to go together in harmony and those that seem discordant. The Greeks, like many other cultures, had developed rules for the playing of the good notes that produced harmonious listening. In modern Western terminology, the notes were collected into scales and octaves. What Pythagoras is said to have discovered is that there is a mathematical relationship between such notes. On the monochord, the relationship could be seen in the length of the string to be played. If the string is played without a stopper, so that its full length vibrates, a certain note is played. If the stopper is then placed halfway, the note that results from only half the string vibrating is an octave higher than the original. If the string were to be doubled in length, it would play a note an octave lower.

In the Western tradition, the separation of these notes is called an octave because there were originally eight notes placed in this interval. However, that is just one possible way of dividing the interval, and different cultures have produced different divisions. For example, in the Chinese tradition there are five notes, while in the Arabic there are 17. Yet while the number of notes placed into an "octave" is variable, all cultures use the same interval. The notes an octave apart are the same note, and they sound the same, just pitched higher or lower. Musicians in all cultures have recognized this interval as a natural phenomenon. What the Pythagoreans showed was that this interval had a mathematical basis, and could be expressed in relation to the length of a monochord string by the ratio 2:1.

Playing two notes an octave apart either together or immediately after one another sounds harmonious. It is said two musical notes are harmonious when they sound pleasing together, as opposed to discordant notes that sound bad together and can make listeners wince or block their ears. Musicians in all cultures had, through trial and error, discovered those notes that seemed to go together well in harmony, simply by listening. The Pythagoreans revealed a mathematical relationship between these harmonious notes.

The modern notes C and G sound pleasing when played close together. If a monochord is set up so the full length of the string plays a C, then to play a G the stopper is moved two-thirds along (a third from the end). The ratio of the lengths of C and G is 3:2. The modern musical term is that these notes are separated by a perfect fifth. Other simple ratios of the string also give rise to harmonious notes. Some notes, however, produce a discordant sound together. Playing a C with an F sharp, an interval known as an augmented fourth with the ratio of 45:32, does not sound good.

The reason for these different sounds, it was discovered centuries later, is the frequency, or number of vibrations per second, at which the strings vibrate. A string of half the length vibrates at twice the frequency of a full string, whereas two-thirds of a string, the perfect fifth ratio, vibrates at one and a half times the speed of the full string. The Pythagoreans concluded that musical harmony was a mathematical property, and occurred when the ratios between the notes were simple.

Pythagoras and his followers believed that the ratios they had discovered using the monochord could not only be applied to other musical instruments, but to the whole of nature. They developed a theory that linked music, mathematics, and the motion of the planets. The Greek view of the universe placed the Earth at the center, with the sun, the moon, the five known planets, and the stars all rotating in fixed crystalline spheres around it. The Pythagoreans applied the same principle of ratios to the

supposed orbits of the planets, and concluded that they had the same properties as a musical scale. The crystal spheres, they said, must make a sound as they moved, and these sounds must be harmonious. Other Greek philosophers, such as Plato, became enamored with this rather beautiful notion of a singing universe, and added to it. The idea of the music of the spheres would remain in fashion for hundreds of years, heavily influencing the study of astronomy.

MEDIEVAL MONKS

The Greeks linked music and mathematics so tightly that the study of music was considered a branch of mathematics, along with arithmetic, geometry, and astronomy. The Greek ideas were, however, almost lost to the Western world after the fall of the Roman Empire. Monks cloistered in monasteries were the only ones with the education and time to translate and copy the surviving writings of the Greeks. Anicis Manlius Severinus Boethius (c. 480–524) translated and copied the ideas of the Pythagoreans. Boethius did more than just copy ancient texts, he also drew together many sources into coherent books. Also, like many other monastic copyists, he edited sections so that they conformed to his own beliefs. Boethius was an excellent translator and copyist, but his grasp of mathematics was poor, and his writings were often hard to follow or misrepresented their sources. However, his copies, mistakes and all, were copied by others and followed blindly, even when they contradicted real-life experiences.

Whatever the limitations of Boethius' copies, they had a profound and lasting effect on Western teaching. Copies of his compilation spread across Europe, and influenced the way music and mathematics were thought of and studied. Music remained a kind of sub-discipline of mathematics, and was taught in European universities as part of the quadrivium ("the four ways"), the same set of subjects that Pythagoras had grouped together: arithmetic, geometry, astronomy, and music.

Under the influence of Boethius, the science of harmonics was the main focus of musical study. The medieval scholars categorized music in three ways, the actual making of music (*musica instrumentalis*), the harmony of the human soul (*musica humana*), and the music of the spheres (*musica mundane*). In this scheme, music was seen as part of the basic nature of the universe, while the playing of music was merely the lowliest part of musical study.

Fixed styles of singing, based heavily on the theories of Pythagoras, became entrenched in the musical practice of the times. Most formal music was made for religious purposes, and just what forms of music were appropriate were written into canon law as the "Ecclesiastical modes." This provided a unifying structure to European music, but also limited innovation. Monophonic chants with one melody line without an accompaniment, where all the singers sing the same words at the same pace, were the standard form. Harmonies were slowly introduced into Gregorian chants by singing the same words at different frequencies at the same time, and these were usually limited to an octave, a fifth, or a fourth apart.

Then a new kind of song type called polyphony, where two or more melodies would be sung or played at the same time, began to take over. The composer had to be careful to make the separate melodies blend together in a harmonious manner, which was rather hard to do. Even harder was trying to write such a song down on paper. Monophonic songs only required the composer to write the pitch or note that was to be sung. A polyphonic song required a method of recording the time each note was to take for all the melodies.

QUANTIFICATION OF MUSIC

In monophonic plainchant, the singers all start together, sing the same words, and stop together, so the level of notation that is required is simple. However, in polyphonic songs, different melodies must be sung, and without a method of knowing the time each note is to take, the whole process slides into anarchy and discordance.

The rise of polyphony, therefore, coincided with advances in musical notation, which also coincided with the development of mathematical notation. Early musical notation developed in the West as a way of fixing songs into a specific form. In the sixth century A.D., Saint Isidore had complained that "[u]nless sounds are remembered by man, they perish," and wondered how reliable were the memories of some singers. However, the early notation could not express the concept of time in music, such as how long to sing a note.

Polyphony required such tools, so musical notation evolved into a way of quantifying time. The melodies a composer had ticking in his head could be transmitted to performers in different places, and even later times, by writing them down using the special code of music.

One of the most interesting developments in musical notation was the ability to note a specific period of silence with rests, which were introduced in the thirteenth century. Around the same time the Hindu-Arabic numerals were being popularized in Europe, and with them the strange symbol for zero. While the ideas of silence and zero may be taken for granted, they were both revolutionary in their respective fields. The musical rests

enabled even more complex melodies to be linked harmonically, and the zero opened up new ways of thinking about numbers in mathematics.

DISCORDANCE OF THE SPHERES

Innovation in music, mathematics, and other fields led to a realization among the scholars of the age that they were doing new things, going beyond the ancients ideas of the Greeks and Romans. In music, the new styles came to be called the *Ars Nova* (The New Art). However, the ideas of the ancients were difficult for many to abandon. The idea of the harmony of the spheres remained a central part of scholarly thinking, partly because it meshed nicely with Christian beliefs and implied an order to creation.

Just as there had been great changes in music and mathematics, the field of astronomy also went through a revolution. In 1543, Nicolas Copernicus (1473–1543) challenged the accepted wisdom of the ancients by publishing a book that suggested the Sun, not the Earth, was at the center of the universe. One scholar who embraced this new idea was Johannes Kepler (1571–1630), who attempted to merge the new Copernican universe with the ancient idea of the harmony of the spheres.

Because the study of astronomy was almost always accompanied with musical study, many important figures in the development of astronomy were also keenly interested in music, and Kepler was no exception. Kepler tried to piece together a model of the universe that used the musical and geometrical ideas of Pythagoras and the new theory of Copernicus, to finally reveal just what notes each planet sang as they moved through the heavens in their crystalline spheres. However, he could just not get the theories to fit together.

After a number of unsuccessful early attempts, Kepler was fortunate enough to inherit a huge collection of accurate planetary observations upon the death of his master, Tycho Brahe (1546–1601). Kepler tried once more, using the observations to guide him in recreating the motion of the planets. He found a regular order in the motion of the planets, but to his great surprise it was an entirely different type of motion than expected, and contradicted all the theories he was attempting to unify. Kepler realized the observation figures he had been given, and those he had made himself, showed that the planets moved in elliptical orbits. While this was not the philosophical order he had been searching for, it was a mathematical order, and he produced equations that predicted the planetary motion with unprecedented accuracy. His insights slowly spread and gained acceptance, shattering the crystalline spheres, and bringing to an end the scientific search for the harmony of the spheres.

WELL-TEMPERED TONES

Music composers grew more and more innovative and daring as the years progressed, experimenting with new styles and ideas such as modulations of scale, just as innovators in other fields experimented with new theories and devices. However, some of the new compositions began to push the limits of the music instruments and tunings of the times. The musical notes that had been evolved from the Pythagorean theories used "just tones," that is, intervals between the notes that were derived from integer ratios. However, this only allowed for perfect tuning in one key at a time. If the composer wanted a piece of music to change keys, then either the musicians had to re-tune on the fly, use different instruments, or sound out of tune.

The problem was that the notes in the just scale were not equally spaced in the octave. When key changes were made, there was a need for new notes to fill in the gaps in the scale. For example, in the diatonic scale the pure (or just) ratios between the notes are either a tone or a semitone apart. The semitones all have the interval 15:16, and the tones are either separated by 8:9 or 9:10, just to make it more complicated. Because the intervals between the notes are different, playing in a different key (which can be thought of as the note started with) means every key has different patterns of intervals between the notes. Also, if the key chosen is too far from the instrument's tuning, so-called "wolf-tones" are heard, where discordant sounds shrill and howl for notes that theoretically should sound harmonious.

Many possible solutions were suggested and tried, and compromise tunings, or temperings, were attempted with various degrees of success. The mean-tone temperament used only major thirds and minor sixth intervals between the notes, effectively averaging out the scale. This meant that all the fifths and fourths that could be played were a little out of tune, but barely. This system worked well for six major and three minor keys, but outside of those, was very discordant.

Eventually the well-tempered tuning was introduced. In this tuning, the interval between each note is made equal, so it is often referred to as the equal temperament. The octave was simply divided into 12 equal parts, giving 12 semitones. In mathematical terms, since the interval ratio of an octave is 2:1, the interval between semitones had to be that number that equals two when multiplied by itself 12 times, which is 1.0595. This meant that changing keys was no longer a problem, as every key has the same interval between notes. However, the beautiful and precise Pythagorean ratios between the notes were now lost, so that all the notes in the octave interval were slightly out of just tuning. The equal temperament sacrificed

accuracy for flexibility, precision for practicality, but in doing so allowed for much more innovation and experimentation in music.

The well, or equal, temperament was suggested as early as 1550, but was slow to be accepted. It became championed by Johann Sebastian Bach (1685–1750), and in 1722 he published a set of 24 preludes and fugues in the 12 major and 12 minor keys of the well tempering, naming them *The Well Tempered Clavier*. Anyone wishing to play these pieces (and Bach's music was quite popular) had to tune their instruments to the equal temperament. Other composers also adopted the new tuning, and slowly it became the standard tuning for Western music, confusing music students ever since by the fact the modern octave now consists of 12 notes.

Real-life Applications

MATHEMATICAL ANALYSIS OF SOUND

A vibrating string produces a fundamental note, the note that is heard when it is played. However, in reality it also produces other sounds called overtones that add complexity to the sound. This is why strings made from different materials will make different sounds even when everything else is equal, or why different kinds of musical instruments can play the same note yet give out a unique tone. However, overtones also add complexity to the analysis of sound. A pure tone, it was discovered, can be represented as a simple wave, but when overtones are introduced, the analysis becomes much more difficult.

The mathematical work of Jean-Baptiste-Joseph Fourier (1765–1830) led to a method for analyzing, and eventually recreating, sound. Fourier discovered that any periodic oscillation, of which sound waves were later shown to be one type, can be broken up into a set of simple sine curves. The sine function is one of the basic functions of trigonometry. A sine curve or wave is defined by the function $y = \sin(x)$, and can be considered as the modeling of a pure tone without overtones. What Fourier's work showed was that complex waves can be thought of as the addition of a number of sine waves of different frequencies and amplitudes. Fourier analysis can take a complex sound wave and break it apart into a collection of simple sine curves. The way the sine curves interact, canceling out in some places and combining in others, means that sound waves, even strange artificial curves such as square or triangular waves, can be broken down and analyzed with some simple mathematics.

The oscilloscope allows regular vibrations, including sound waves, to be displayed in real-time. An oscilloscope, in essence, draws a graph on the screen, displaying the signal as a waveform, which can be broken down into a sum of sine curves. Oscilloscopes are used by scientists, television and automotive repair technicians, in medical research, and to measure and analyze diverse phenomena from stress in buildings and brain waves, and, of course, sound waves.

If sound can be deconstructed into simple sine curves, then simple sine waves can be generated, then added together to reconstruct, or synthesize, any sound. This is the principle of modern electronic music synthesizers. Theoretically, the quality of the sound these instruments can produce depends only on the accuracy of the original analysis of the sound to be reconstructed and the quality of the sine waves that can be produced.

ELECTRONIC INSTRUMENTS

In the last 100 years, the electrification of music has changed the way music is produced and listened to. Electric instruments were first introduced as a means of making a louder sound. For example, the Hawaiian-style, or slide, guitar was a popular instrument in the 1930s, but because it was played horizontally it projected most of its sound upwards, rather than toward the audience, limiting the size of the audience it could reach. In the 1930s, the Gibson company successfully placed an electronic amplifier inside an otherwise acoustic Hawaiian guitar, and the electric guitar was born.

Distortions introduced in the amplification process limited the volume of early instruments and their acceptance. However, by the 1950s electric guitars produced cleaner, louder sound, and could sustain notes for a longer time than any acoustic model. The rising popularity of the electric guitar ushered in a whole new range of sound, such as the electronic manipulations of the sound, using fuzz-boxes, wah-wah peddles, and many other devices. The loudness of the electric guitar had to be matched with louder supporting instruments, and so all the instruments had to be wired up and their sound amplified.

Early electronic instruments still relied on a natural vibrator, such as string or drum skin, even though the final sound might bear little resemblance to the natural sound of the vibrating element. However, since the mathematical analysis of sound had shown that sound waves could be thought of as the sum of simple oscillations, then it must be possible to build sounds from a simple electronic source, and so the synthesizer was born. The Hammond electric organ, which debuted in 1935, can be considered the first of the electronic synthesizers, although the technology it used relied on many moving

parts. Later electronic keyboards use a myriad of mathematical algorithms to reproduce the sounds and rhythms of other musical instruments. The synthesizer contains an oscillator to produce the basic frequency, which is then amplified, mixed with other frequencies to add richness to the sound, filtered to remove unwanted noises, and can then be submitted to a host of other modulations to give the final desired sound. Early synthesizers sounded artificial, with sounds that mimicked actual instruments, but did not sound as natural. Better mathematical analysis of sound, combined with better programming and electronics, has produced much richer sounding electronic instruments, with the ability to not just sound like an genre of instruments but an actual specific instrument, such as a particular Steinway grand piano or an individual Stradivarius violin.

Personal computers (PCs) have added another dimension to electronic music, with many add-on components and programs that turn the PC into a recording studio. A single instrument can, through the manipulation of computer algorithms, produce the tracks for a complete band or orchestra. While early programs were little more than novelties, producing bleeps and beats, it has become possible to record professional quality music on a PC, and a number of professional musicians have done so. The flexibility and sophistication of computer-generated music has even allowed musicians to experiment with scales other than the well tempered scale, and many alternate tunings that use the just or Pythagorean scales have again become popular, as their limitations can be overcome electronically.

ACOUSTIC DESIGN

Making sound louder by electronic amplification is one method of making sounds easier to hear. However, for some styles of music, amplification is not an option, such as classical or acoustic concerts. Either the number of people who can hear the music will be very limited, or some other method of boosting the strength of sound to help it arrive to a listener's ear must be found. Acoustic design is a branch of architecture that attempts to build concert halls, sound stages, and auditoriums in such a manner as to maximize the amount of sound that reaches the audience.

Sound quickly fades in strength as one moves from the source, in proportion to the inverse square law. So if one moves twice as far away from a sound, one hears a quarter of the original strength. However, sound can be reflected with the right surface. Ancient Greek amphitheaters used hard rocks in large amounts to help large audiences hear the voices of actors, using clever angles to get the most reflection and the best focus on the listeners as they could. The designs for modern concert halls and stadiums are computer modeled to maximize the sound reflection, while removing unwanted reverberations. Geometry plays a key role in such design, as well as the reflective and absorbent qualities of the materials used. By integrating different materials with careful design, the sound of a room can be crafted to give a warm, rich sound or a clean, intimate sound, or many other desired results.

For many buildings, it is sound reduction or diffusion that is important. By using materials that absorb sound and angles that diffuse rather than concentrate, architects can design rooms where sound does not travel clearly, for example, to avoid secrets being overheard. Sound insulation and soundproofing can effectively isolate the sound of a room from rooms or areas close by.

DIGITAL MUSIC

A digital revolution in the late twentieth century changed the way music is recorded, stored, and listened to. Early music recording and storage devices were analog devices, such as LPs (long-playing records) and audiotape. Analog means a continuous property that varies, such as the bumps on a wax cylinder, the groove of a record, or the magnetic alignment of grains on an audiotape. Analog recordings are very susceptible to errors, such as a scratch on a record, or being dropped or jolted, and are inefficient, needing a large surface to record small amounts of information. Digital music was introduced as a way to eliminate errors, and as a more compact medium. The compact disc is much smaller than an old LP, yet can store almost twice as much music, while a computer hard drive can have an entire music collection stored in digital form.

Digital simply means the information is recorded as binary numbers, ones and zeros. The sound to be encoded is sampled a large number of times each second, and a value is given for the wave height at that moment. Compact discs (CDs) were designed for a sample rate of 44,100 times per second (44.1 kHz). Going from analog sound to digital data is an inexact process, and even with a high sample rate some information will be lost or simplified, because the number of samples is limited, and the acceptable values for the wave height are also restricted.

Like building a staircase, sampling sound has to consider the distance between the steps (the sample rate) and the height difference between the steps (the allowed values of the sampling, referred to as the bit depth). The wider the steps, the harder it is to walk up them, and by analogy, the sound is poorer in quality, because the sampling is too far between steps. The taller the steps, the harder they are to walk up, and by analogy, the less accurate the sample

estimations are. The solution is to make short, low steps, to sample frequently, and to have many small increments in the values the estimates can take. However, there is a practical limit to how small these steps can be made and still be usable. In the case of digital music, the limit is how much storage space the information requires.

COMPRESSING MUSIC

CDs use comparatively large amounts of space to store their digitally encoded information. Binary information on the CD is stored in groups of 16 ones and zeros (16-bit samples). There are 8 bits per byte, so 44,100 16-bit samples per second equals 88,200 bytes per second. That's for one channel; twice as much is needed for a stereo recording. Consequently, 176,400 bytes per second is about 10 megabtyes (10,584,000 bytes) per minute for stereo recording.

One method to reduce storage size is to take less-frequent samples (lower sample rate) or less accurate values (less bit-depth), but this leads to a poorer sound quality. So compressing the data is the preferred option, used in formats like the mp3. A range of mathematical topics contributes to the field of data compression, including algebra, statistics, graph theory, calculus, Fourier analysis, and fractals. Basic compression takes redundant information that can be simplified to the point where it takes less binary numbers to encode.

Another compression technique relies on the fact that the human ear is an imperfect listening device. The ear cannot detect all the complexities of sound, and will often be unable to tell the difference between a complex noise and a simplified substitute. Mathematical models of the human ear have helped define methods for dramatically reducing sound files in this manner, essentially tricking the ears of those who listen.

ERROR CORRECTION

Whenever a compact disc skips or an mp3 file gives out an unexpected sharp squawk it is because of an error. These errors can be from physical substances, such as dust, dirt, or scratches, from jolting a music player, or from recording or manufacturing mistakes. Because errors are so common, music players must use error-correction algorithms to fill in missing information where possible.

The mathematics for correcting these errors were discovered many years before any use for them existed. The work of Claude Shannon (1916–2001), Warren Weaver (1894–1978), and others on information theory (IT) showed that information transmission errors could be corrected by mathematical algorithms, and later work by other mathematicians and engineers provided working applications. The basic ideas of error correction are to add extra, or redundant, information to the signal so that there is a better chance of one of the bits of information being read correctly, and the information encoded in such a way that it can self-check as it goes along, so that it literally knows what to expect next. The mathematics of error correction is also used in linguistics, psychology, cryptography, and the elimination of noise in the transmission of data.

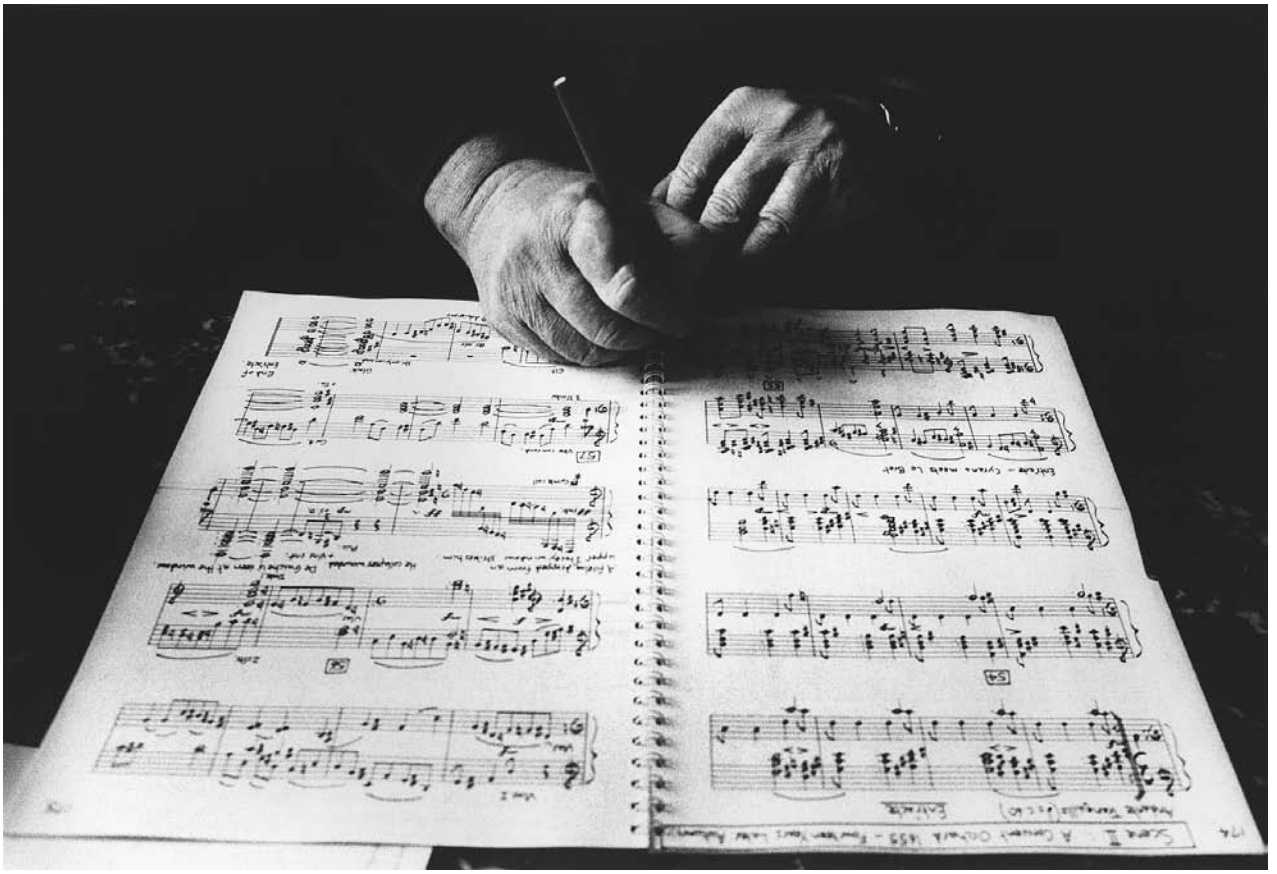
USING RANDOMNESS

Mathematical ideas have often been incorporated into composition techniques. Chance music, also called algorithmic, aletoric, random, or stochastic music, allows for unexpected structures to be introduced as defined by a set of rules, for example, by tossing a coin. Some chance pieces have strict rules that provide a high degree of structure, while others are so flexible in the choices that can be made as to be almost unpredictable.

Wolfgang Amadeus Mozart (1756–1791) created a musical game, *Musikalisches Wurfelspiel* (Dice Music), in 1787, where a minuet is formed by rolling dice, which determines the order of pre-written measures of music, sort of like cutting and pasting. The background music for the 1938 film, *Alexander Nevsky*, composed by Sergei Prokofiev (1891–1953), used the landscape in the film as a pattern for the notes. John Cage's (1912–1992) *Atlas Eclipticalis* (1961) was composed by placing the score over star-charts, with the position and brightness of the stars visible through the paper determining the notes, while his Reunion (1968) is performed by playing chess on a chessboard equipped with photo-receptors, each move determining the series of sounds to be played. Many of Cage's pieces allow for great freedom of interpretation, including how many instruments, or what instruments, are used, and the pitch, duration, intonation, and loudness of the notes to be played. Another example is the composition of Brian Eno's (1948–) *Music for Airports* (1978), where pieces of prerecorded audio tape were cut up and several played simultaneously in loops. These laborious cut-and-paste techniques of composition were later revolutionized by personal computers.

COMPUTER-GENERATED MUSIC

Human composers may incorporate random or chance elements in their work, but wisely tend to let such processes guide them rather than dictate the final results. Getting a computer to compose music means letting go of that human input, and gives rise to many problems. A number of attempts to have a computer compose music



Musical scores convey information with almost mathematical elegance. JOHN GARRETT/CORBIS.

were made as early as the 1950s, with limited success. Sometimes the resulting pieces had snatches of tuneful music, but were overall disappointing, while other attempts produced listenable, but extremely simple tunes. In 1959 the ILLIAC computer, an early supercomputer, was programmed with the rules of composition that had been written during the period of Baroque music. The *Illiac Suite for String Quartet* was performed successfully, but the composition was not considered to be of high quality.

The problems with computer composition were that when there are few rules for generating the notes, the resulting tunes are too random in structure, resembling white noise. The music generated has no direction or coherence. If the rules are stricter, then some coherence can be found in some tunes, because rather than white noise the use of strict algorithms often produces brown noise. Brown noise, named after Robert Brown (1773–1858), refers to a type of process that can be thought of as a random walk. The starting point can be set, but after that the walker may end up moving forward or back, to the sides, or some combination. Over time, the walker will have progressed somewhere and how they will get there cannot be determined. Not surprisingly, brown-noise computer compositions sometimes have a direction, of sorts, but can be rambling and dull.

Fractal mathematics, popularized by Benoit Mandelbrot (1924–) in the 1970s, offer a different type of structure for computer-generated music. Fractals are curves, surfaces, and objects that have non-integer dimensions. A point has a dimension of zero, a line a dimension of one, a square two, and a cube three. However, fractals have dimensions that lie between these; for example, a geometric construction called Koch's cube has a dimension of about 1.26. Surprisingly, a number of natural phenomena displays fractal properties, such as clouds, coastlines, landscapes, plants, and many more. Fractional dimensions produce some interesting patterns that usually have a degree of self-similarity in them; that is, the large-scale pattern resembles smaller structures within the pattern, which in turn resemble smaller structures within themselves. For example, a tree has a growth pattern that resembles that of a single branch, and within a branch there can be smaller branches with the same pattern.

Compositions that use fractal formulas sound more coherent that other types of computer-generated chance music. The property of self-similarity means that fractal compositions repeat themes in complex ways, which closely mimics a common property of human composition. Without knowing it, composers like Prokofiev and Cage had already used fractals in their music by using landscapes and star patterns to determine notes. However, most computer-generated music still sounds aimless and flat in comparison with human compositions. While many musicians have embraced the flexibility and inspiration that computer generation can give to a composition, computers are in no danger of taking over the writing of music just yet. There is still something human choices can give to music that cannot be fully simulated.

Randomness was originally seen as a negative when it entered music as unexpected or unwanted noise, yet when harnessed in the right manner it has produced many innovative and important pieces of music. Randomness has also entered into the way music is listened to, from CD-shuffling stereos to mp3 players with a random song selection, the old structures of albums and playlists are often sidestepped.

FREQUENCY OF CONCERT A

Sometimes mathematics and music have come into conflict, such as the long debates over the correct frequency of the notes in the Western scale. Orchestras and many other musicians often tune their instruments to the note known as concert A, and from that all the other notes in the scale are then set by their musical intervals from that frequency. However, the choice of this frequency is arbitrary. At first, Western music had no standard frequency for concert A, as there was very little communication across medieval Europe. Different regions sang and performed with their own pitches, because they had their own frequencies for the same notes. However, as contact between musicians increased across Europe, a rough standard was introduced and in the eighteenth century, concert A, as estimated by music historians, was about 420–425 Hz (Hertz, or cycles per second).

Once sound frequencies became better understood, and methods of measuring frequency were available, there were attempts to introduce a more specific and universal standard for concert A, although national pride and politics got in the way. The French and English set different frequencies, of 435 Hz (cycles per second) for the French and 439 Hz for the English.

Then in 1939, an international standard of 440 Hz was introduced, but against the will of a mathematical lobby that wished concert A to be set at 426.7 Hz, so that middle C would be at 256 Hz. This was called the philosophical pitch, as 256 is 2^8 (two to the power of eight, or two multiplied by itself eight times), and so seemed to the mathematicians a more formal, even Pythagorean, derivation of the note. The musicians, however, did not want such a dramatic change to the pitch of the music they played, as such a low number for concert A would have altered the sound of all existing music.

MATH-ROCK

Most musicians do not consider the mathematics that lie behind their music. Music can be composed and performed extremely well without any mathematical input from those involved. However, with the introduction of electronic instruments, it has become easier to introduce mathematical concepts into music. There is even a genre of rock music that calls itself math-rock, and is categorized by the creative use of time signatures.

A time or meter signature can be thought of as the number of beats in a measure of music, or in basic rhythmic terms, the number of drum beats in a set period of time. Normally, all the instruments in a piece of music will play in the same time, as this makes it easy to keep together, and usually sounds better. If one instrument plays in a different meter than the others, the result is usually unpleasant to the listeners. In math-rock, however, the musicians play in different meters on purpose. For example, in the Frank Zappa (1940–1993) instrumental *Toads of the Short Forest* from the 1970 album "Weasels Ripped My Flesh" there are two drummers, one playing in 7/8 time, while the other plays in 3/4. At the same time the organist plays in 5/8, creating an effect known as polyrhythm, or polymeter. The roots of polyrhythmic music go back to Indian and African music, as well as Latin music.

Music performance and composition are art forms, and many have called mathematics at its highest levels more art than science. Yet, when the basics of mathematics or music are learned, they both must start with simple rules and learned by rote, memorizing the building blocks of the subject until they become second nature. As learning progresses, the rules become more complex, and the effort needed to master them increases. For some people the effort is too great, or the rules too complex. Only a few people master a branch of mathematics or a genre of music, and at the highest levels the rules do not seem to be so important. They are still there, underpinning everything, but can be used in new ways, or stretched, or combined with unexpected results. For those people on the outside looking in, these highest workings of music and mathematics may be fascinating,

Key Terms

Analogue: A continuously variable medium, for use as a method of storing, processing, or transmitting information.

Frequency: Number of times that a repeated event occurs in a given time period, usually one second.

Scale: The ratio of the size of an object to the size of its representation.

spellbinding, even beautiful, but are strange and unexplainable; they are to be enjoyed, but never fully understood.

Potential Applications

Mathematics and music have become more entwined than ever before. The ways music is made, produced, transmitted, and listened to all rely heavily on mathematics, and many practical applications in music have come from abstract mathematical concepts. The shift to digital formats for music has been accompanied by continuing work in compression, error correction, and improving quality. Work in the field of music has produced mathematical tools and applications for other areas, and will continue to do so. In turn, mathematical ideas in other fields have successfully been transposed into musical applications. The experimental ethos that is at the heart of musical expression also exists in the field of musical instrumentation and engineering, and new devices are constantly being created.

Where to Learn More

Books

Garland, Trudi Hammel, and Charity Vaughan Kahn. *Math and Music: Harmonious Connections*. Palo Alto, CA: Dale Seymour Publications, 1995.

Helmholtz, Hermann. *The Sensations of Tone*. Mineola, NY: Dover Publications, 1954.

Johnston, Ian. *Measured Tones: The Interplay of Physics and Music*, 2nd Ed. Bristol, England: Institute of Physics Publishing, 2002.

Roederer, Juan G. *The Physics and Psychophysics of Music: An Introduction*, Third Edition. New York: Springer Publishing, 2001.

Web sites

Mozart's Musikalisches Wurfelspiel (Musical Dice Game) November 1995. <http://sunsite.univie.ac.at/Mozart/dice> (May 25, 2005).

Music and Mathematics: Dave Benson. 2003. <http://www.math.uga.edu/~djb/html/math-music.html> (May 25, 2005).

Nature and Numbers

Overview

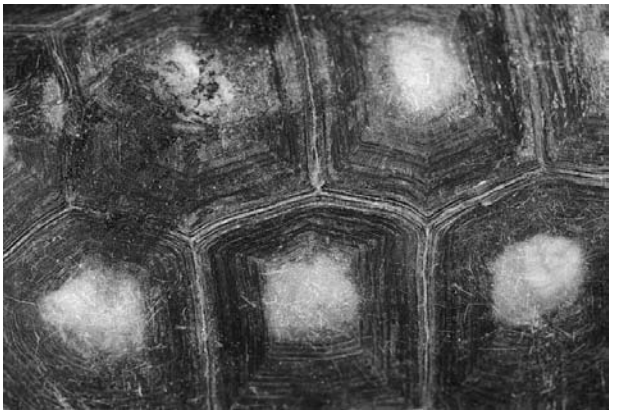
Scientists depend on numbers and mathematics to explain many of the patterns that they observe in nature. For example, scientists find the same numbers repeated in many different places in nature. Fibonacci numbers and the Golden section are found in the geometry of mollusk shells and in the petal and leaf patterns of plants. Scientists also use special mathematical formulas called models to predict patterns that occur in nature. Mathematical models are used to explain how horses move and where phytoplankton grow most quickly in the ocean. New fields of mathematics have also resulted in insights into nature. Fractals are mathematical rules that can be used to represent complex shapes in nature and knot theory has the potential for providing an improved understanding of viral infections. Both numbers themselves and a variety of mathematical techniques greatly improve our understanding of the natural world.

Real-life Applications

FIBONACCI NUMBERS AND THE GOLDEN RATIO

Fibonacci (1175–1250) was an Italian mathematician credited with introducing the decimal system to Europe. In 1225 he took part in a tournament to solve a mathematical problem: a pair of rabbits produces another pair of rabbits every month, but it takes two months before any pair can produce their first offspring. Assuming no rabbits die, at the end of a year, how many rabbits will there be? Fibonacci answered the question by writing out a series of numbers. The first month there is one pair. Since the pair cannot reproduce until after the second month, the second month there is still just one pair. The third month, the pair reproduces so there are two pairs. The fourth month, the original pair reproduces, but the younger pair is still too young to reproduce; there are three pairs of rabbits. Continuing with this logic, Fibonacci showed that at the end of 12 months, there would be 144 rabbits. The series of numbers that answered the question is 1, 1, 2, 3, 5, 8, 13, 21, 34, 55, 89, 144 . . . Fibonacci noticed that this series could be extended infinitely by adding the previous two numbers to produce the next number in the series.

The Fibonacci numbers are repeatedly found in nature, particularly in plants. The number of petals on most flowers and leaves on many plants are Fibonacci numbers. For example, calla lilies have one petal, irises and lilies have three petals, buttercups, roses and columbine have five petals, delphiniums have eight petals, some daisies and marigolds have 13 petals. The numbers of leaves on only a few plants are not Fibonacci numbers: consider how rare



Note the geometric detail of a tortoise shell. ROYALTY-FREE/CORBIS.

four leaf clovers are. The center of many flowers, like sunflowers, is called the seedhead and it consists of seeds packed together in spirals. These spirals are organized in both clockwise and counterclockwise directions and they are interlocking. On a flower's seedhead the numbers of spirals in both directions are also Fibonacci numbers. For example, the seedheads of coneflowers have 55 clockwise spirals and 34 counterclockwise spirals; sunflowers have 89 clockwise and 55 counterclockwise. It turns out that this arrangement of seeds allows for optimal packing: seeds are not crowded and the most seeds possible can fit on the seedhead. The same pattern of interlocking spirals also appears on cauliflower heads, broccoli heads, pinecones and pineapples. The numbers of spirals on these structures are also always Fibonacci numbers. Additionally, the branching patterns of many plants can be expressed using Fibonacci numbers. The patterns of leaf growth around plant stems, called phyllotaxis, uses Fibonacci numbers.

Dividing every number of the Fibonacci series by its previous number results in another series: 1, 2, 1.5, 1.66, 1.6, 1.625, 1.61538 . . . , which eventually converges to 1.618034. This number is called the Golden ratio, the Golden section or the Golden number. It is often represented by the Greek letter Phi, ϕ. The part of the number to the right of the decimal, 0.618034 is called phi (with a lowercase p) and is equal to $1/\phi$. Both Phi and phi are observed in many places in nature. For example, many spiral mollusk shells incorporate Phi in their growth pattern. The nautilus shell is created as the mollusk adds larger and larger chambers to the outside of its shell. Drawing a line from the center of the spiral to the outside, the ratio of the distance between one turn of the spiral and the next turn of the spiral along that line is Phi. The DNA molecule, which contains the genetic information for all organisms, has the shape of a double helix. Each turn of the helix measures 34 angstroms in length. The width of the DNA molecule is 21 angstroms. These numbers are successive Fibonacci numbers and their ratio is close to the Golden ratio. The Golden ratio is observed in ratio of the width to the height of a dolphin's fin. The ratio of the length of the bones at the tips of the human fingers to the length of the bones between the first and second knuckle is the Golden ratio. Similarly, the ratio of the length of bones between the first and second knuckle to the length of the bone between the second knuckle and the hand is also the Golden ratio. The Golden ratio also turns up in the ratio of the distance between the top of the human head and the navel and the distance between the navel and the feet. In fact, some consider the Golden ratio fundamental to art, architecture and music.

MATHEMATICAL MODELING OF NATURE

A group of equations that are combined together in order to make a prediction is called a mathematical model. Scientists and mathematicians often use mathematical models to better understand their observations of nature. Mathematical models explain a variety of processes in the world, from how horses move to where phytoplankton are found in the ocean.

Biomechanics is a field of research that investigates how the materials in living organisms behave and mathematical models are fundamental to this research. For example, biomechanists have developed a mathematical model that describes the stresses and strains on the bones and muscles in horses' legs. These models lead to predictions of how horses move. At rest, the stress on a horse's bones is about 25% of the stress it would take to break the bones. When a horse starts walking the stress increases and it continues to increase as the horse walks more quickly. At the point where the stress reaches 30% of the breaking point, the horse switches to a trot, relieving the stress on the bones. Again the stress increases as the horse trots more quickly. The model predicts that at the point when the stress is about 30% of the breaking point, the horse should switch its behavior. Indeed at the speed where the stress on the bones is 30% of the breaking point, the horse switches its motion and begins to canter and the stress decreases. Finally, when the canter speed increases to the point that stress reaches 30% the model suggests another change in behavior. This speed is, in fact, the speed at which the horse begins to gallop decreasing the stress once again.

Phytoplankton are microscopic plants that live in the ocean. They perform photosynthesis, which is responsible for producing most of the oxygen on Earth and for removing a great deal of the greenhouse gas, carbon dioxide, from the atmosphere. In addition, phytoplankton are the base of the oceanic food chain, so where phytoplankton are, there are usually fish to be caught. Phytoplankton

Key Terms

Fibonacci numbers: The numbers in the series, 1, 1, 2, 3, 5, 8, 13, 21, 34, 55, 89, 144 . . . , which are formed by adding the two previous numbers together.

Fractal: A self-similar shape that is repeated over and over to form a complex shape.

Golden ratio: The number 1.61538 that is found in many places in nature.

Knot theory: A branch of mathematics that studies the way that knots are formed.

grow by dividing into two individuals, so when they grow faster, they are found in higher concentration. Phytoplankton growth requires both light and nutrients and it usually occurs more quickly when water temperature is warmer. Given information about light levels and nutrient concentrations, scientists construct mathematical model that predict where phytoplankton are found and in what concentration they will be found throughout the ocean. In fact, scientist can use their models with satellite imagery of the ocean to predict concentrations of phytoplankton throughout the oceans. These predictions help estimate the effects of carbon dioxide on the planet and help regulators develop fisheries policy.

Mathematical modeling is not limited to these two examples. Practically any observation or a pattern in nature, such as describing how leopards get their spots, why monkeys could never grow as large as King Kong, where hurricanes are likely to make landfall and even the height of the tides, can be better understood by developing mathematical models.

USING FRACTALS TO REPRESENT NATURE

In the 1970s a new branch of mathematics emerged that demonstrates how complex shapes in nature result from the repetition of the same pattern over and over. These repeating patterns, or self-similar shapes, are called fractals. Notice that a branch of a tree has roughly the same shape as the tree itself. Likewise, a piece of a cloud looks similar to an entire cloud. These complicated shapes are built by writing computer codes with very simple rules, or fractals, and then letting the computer repeat the rules over and over. The result is computer simulations of shapes that look like forms from nature: the complicated branches of trees, the clouds in the sky or a sprig of lilacs on a branch. Fractals have also been used to help forecast weather patterns. Fractals have helped mathematicians and biologists understand that complex-looking forms in nature may not really be that complicated. Instead these shapes are intricate repetitions of simple patterns.

Potential Applications

SPECIFY APPLICATION USING ALPHABETIZABLE TITLE

About one hundred and fifty years ago, mathematicians developed a knot theory, which describes the way that strings are formed into knots using mathematics. This allowed them to compare different knots to see whether they were similar or not. The mathematics of knot diagrams was worked out in the early 1920s using a new form of algebra. Eventually computer codes were developed to study different knots.

Recently knot theory has been applied to the study of viruses. Viruses are really just pieces of DNA or RNA, which can be thought of as long strands that can be made into knots. The virus inserts itself into the host's DNA causing it to coil, or knot, in ways that are very different from the original piece of DNA. Studying these patterns will, hopefully, bring insight to the fight against diseases that are caused by viruses.

Where to Learn More

Books

Devlin, Keith. *Life By the Numbers.* New York: John Wiley & Sons, Inc., 1998.

Gardner, Robert, and Edward A. Shore. *Math in Science and Nature.* New York: Franklin Watts, 1994.

Web sites

Britton, Jill. "Fibonacci Numbers in Nature." <http://ccins.camosun.bc.ca/~jbritton/fibslide/jbfibslide.htm> (September 25, 2003).

"Fibonacci Numbers and the Golden Section" *University of Surrey* <http://www.mcs.surrey.ac.uk/Personal/R.Knott/Fibonacci/fib.html> (November 1, 2004).

"What is the Fibonacci Sequence" <http://techcenter.davidson.k12.nc.us/Group2/main.htm> (November 5, 2004).

"SeaWiFS Project" *NASA Goddard Space Flight Center* <http://seawifs.gsfc.nasa.gov/SEAWIFS.html> (November 5, 2004).

Negative Numbers

Overview

Negative numbers are like mirror images of positive numbers. Any positive number can be made into a negative number by attaching a negative sign to it: for 5 there is −5, for 0.133 there is −0.133, and so on.

While negative numbers may at first seem intuitively odd, they actually have a prominent role in everyday life.

Fundamental Mathematical Concepts and Terms

When representing physical quantities with numbers, negative numbers allow the "bracketing" of a zero value. If that zero value has some meaning, then the positive and negative numbers that accrue from that point also have meaning.

For example, negative numbers allow benchmarks of physical behavior to be established. A fundamental example is the physical liquid-to-solid change that occurs in water at 32 degrees Fahrenheit (0 degrees Celsius). On several scales, measuring temperature negative numbers because the numbers characterize temperature states (a measure of molecular activity) below the zero point. Absolute zero on temperature scales is characterized by negative numbers on the Fahrenheit, Celsius, and Kelvin temperature scales. Absolute zero—0 Kelvin, −459.67° Fahrenheit, or −273.15° Celsius—is the minimum possible temperature: the state in which all molecular motion of the particles in a substance has ceased.

A Brief History of Negative Numbers

The conceptual roadblock against numbers less than zero dates back thousands of years. The ancient Chinese calculated numerical solutions using colored rods. Their use of red rods for positive quantities and black rods for negative quantities is opposite our present-day accounting standard (where "in the black" means having a *positive* bottom line). Although they accepted that numbers could decrease the value of a quantity, the idea that a final solution could be negative was unacceptable. Instead, they would rearrange the problem so that a positive number answer was obtained.

This line of thinking persists. For example, a section of a 2002 United States government income tax document contained the following section: "If line 61 is more than line 54, subtract line 54 from line 61. This is the positive amount you OVERPAID. If line 54 is more than line 61,

subtract line 61 from line 54. This is the positive amount you OWE." In both cases, the solution is presented as a positive, rather than a negative.

The earliest known records of the acceptance of zero as an actual value and the related concept of negative numbers date from India around 600 A.D. Several centuries later, the Indian scholar Brahmagupta routinely used negative numbers and even derived mathematical rules to deal with such numbers.

Yet, even into the 1500s, European mathematicians argued that negative numbers could not exist, since zero signified nothing, and "it is impossible for anything to be less than nothing." In *The Principles of Algebra*, published in 1796, William Fredn stated that "to attempt to take [a number] away from a number less than itself is ridiculous."

By the nineteenth century, however, negative numbers had become accepted as a valid part of numbering systems.

Real-life Applications

TEMPERATURE MEASUREMENT

Whether the thermometer scale is in degrees Fahrenheit (°F) or Celsius (°C) or both, a range of positive and negative numbers brackets the zero mark.

In degrees Celsius, the zero value is arbitrarily assigned to that temperature at which water changes its chemical structure from a liquid to become a solid. Because the temperature of the air can become even colder, it is necessary to have a number scale to relate this degree of coldness in terms that are rational and intuitively understandable.

In a typical thermometer, the expansion and contraction of mercury or alcohol liquid in a column is indicated by the temperature scale. Thus, the mercury column will be shorter (compressed or contracted) on a −10°F-day than the length of the column on a 10°F-day.

The change in temperature that will occur during the course of a day is also described by negative numbers. As the day cools into evening, the temperature will decrease and the mercury column will drop (compress). The temperature decrease is represented by a negative number.

The Kelvin scale developed in the mid-1800s by the British scientist Lord Kelvin. The zero point of the Kelvin scale corresponds to absolute zero. This temperature, which represents the minimum molecular motion, is the equivalent of −273.16°C.

ACCOUNTING PRACTICE

The fundamentally important economic practice of bookkeeping, exemplified in balancing income and expenditures or in the completion of an income tax return, requires negative numbers. Depending on the accounting system used, negative numbers can represent a debt (e.g., −$20 representing a sum owed and to be deducted from an account), an expenditure (e.g., −$20 to be deducted from an account for a purchase, or a remaining balance (e.g., −$20 as a balance when adding together all income and expeditures where expeditures exceed income by $20.

The countdown clock is stopped at the T-minus five minute mark (a negative number) while the Space Shuttle Columbia sits on Launch Pad 39-B on Oct. 15, 1995 at Kennedy Space Center. NASA often uses negative number during countdowns to launch. AP/WIDE WORLD PHOTOS. REPRODUCED BY PERMISSION.

THE MATHEMATICS OF BOOKKEEPING

Bookkeeping describes the process where the amount of money coming in (income) and the amount of money being spent (expenses) are itemized in an arrangement that clarifies which side of the "ledger" is greater.

In a business or a home budget, the ultimate aim is to have a greater income than expense. This remainder can be used for investment or pleasure.

However, a harsh reality can be when expenses exceed income. An annual income of $75,000 and expenses of $85,000 will produce a final tally of 75,000 − 85,000 = −10,000. In modern-day bookkeeping parlance, this negative tally is "in the red." This is also known as a deficit.

Similarly, income tax calculations itemize all the sources of income and expense to arrive at either a positive number, which represents the refund payable by the government, or a negative number, which is the amount that the person or business filing the tax return owes the government.

Such negative numbers are not usually a cause for celebration.

However, a business owner can be heartened if the deficit has decreased from that of the previous year. For example, a deficit for one tax (fiscal) year of $50,000 (−50,000) followed by a deficit of $10,000 (−10,000) in the following fiscal year would be hailed as a tangible indication that the business is edging toward profitability (a positive number). So, even though the "bottom line" is a negative number, the trend is encouraging.

A bank account is another example of bookkeeping. An algorithm (a set of instructions that denote a method for accomplishment of a task) applied when every deposit or withdraw of funds are made keeps a tally of the money remaining in the account. A negative number, which can be represented by the bracketing of the number, indicates that the account is "overdrawn"; more money has been taken from the account than was actually in the account. The imbalance must be corrected by depositing more money to at least bring the account balance to zero.

SPORTS

In the realm of sports, various statistics and scoring systems are rooted in negative numbers.

Negative numbers are in integral part of a number of sports. In golf, the aim on each hole of the 18 holes that make up a standard course is to hit the ball from the starting area (the tee) into a hole in a determined number of shots. The number of shots representing normal or "par" performance varies from three to five (and, in rare cases, six) depending on the length of the hole. The cumulative score represents what is called "par" for the course.

The vast majority of golfers will never achieve par. Their scores will be greater than the ideal score. For example, on a par 72 score, a competent golfer may routinely shoot 85. This can also be described as shooting 13 over par, or +13.

Professional golfers are able to routinely shoot the par score and even lower, as for example a score of 65 on a par 72 score. This score can be described as a negative number, in this case 7 under par, or −7.

While a negative number is desirable in golf, it is undesirable in football. A team's movement on the football field from their end to the opposition's end is measured in yards. Positive yardage indicates forward movement. Negative yardage is indicative of backwards movement. If the quarterback or running back is tackled behind the place where the play began, the effort results in a negative number of yards. This makes the team's task of reaching the opposition's end of the field all the harder.

Negative numbers are encountered in sporting events where time is measured. Running events on the track or longer distances run on roads are two examples. Often, an athlete will gauge his or her performance by splitting the event into two or more equal distances (called splits) and timing each phase of the race (split times). A desirable goal is to achieve what are called negative splits, where the time to complete the later stage(s) of the race is less than the time spent in the earlier distance. Negative split times can be a way to the winner's podium.

On the track, negative numbers are part of the blindingly quick sprints. A worldclass sprinter can run 100 meters in under ten seconds. Often the field of runners will cross the finish line within a fraction of a second of one another. To sort out their finishing places, cameras positioned at the start and finish lines accurately record the respective times. The winner's time can then become the benchmark, denoted as zero, on which the other times are compared by means of negative numbers. Thus, if the second and third place finishers were 1/100 and 1/10 of a second slower, respectively, their times will be displayed as −0.01 and −0.1. This use of negative numbers permits a rapid assessment of the race's outcome.

FLOOD CONTROL

Land located next to a water body that can rapidly increase in volume can be flooded if the increased volume cannot be accommodated. The flooding potential of the watercourse can be measured by recording the level of the water above or below the flood stage. A positive number, which is above the flood stage, indicates that flooding is occurring. A negative number indicates that the water level has not reached the danger zone.

As with temperature, monitoring the progression of the number change over time reveals the trend, and so can guide subsequent actions. For example, if flood control officials note that the flood measurements are negative numbers and these are increasing with time, they can be confident that the flood danger is past and the river's

level is settling back to normal. Flood control procedures can be eased or cancelled.

BUILDINGS

The floors in a multi-story building are denoted by numbers. Sometimes a building has levels below the ground, typically parking space, as well as floors towering overhead. The ground floor forms the demarcation between the above- and below-ground floors, in essentially the same way that zero marks the positive and negative number series on a numerical scale.

The real-life demarcation is readily evident when riding down in an elevator. Beginning on a floor (the fifth, for example), the floor indictor could display 5, 4, 3, 2, 1, 0, −1, −2. The latter two are the underground levels. The practice of which floor to assign zero or ground level often varies from country to country.

Where to Learn More

Books

Pickover, Clifford A. *Wonders of Numbers: Adventures in Mathematics, Mind, and Meaning.* New York: Thomas Dunne Books, 2004.

Schwarz, Alan. *The Numbers Game: Baseball's Lifelong Fascination with Statistics.* New York: PowerKids Press, 2004.

Strazzbosco, John. *Extreme Temperatures: Learning About Positive and Negative Numbers.* New York: PowerKids Press, 2004.

Web sites

Purplemath. "Negative Numbers Review I." <http://www.purplemath.com/modules/negative.htm> (February 11, 2005).

British Broadcasting Corporation. "Skillswise factsheet: negative numbers – practical examples." <http://www.bbc.co.uk/skillswise/numbers/wholenumbers/whatarenumbers/negativenumbers/> (February 11, 2005).

Number Theory

Overview

Number theory is the study of numbers, in particular integers. Integers are the positive and negative whole numbers: . . . −3, −2, 1, 0, 1, 2, 3 . . . Number theory was once considered a branch of pure mathematics, which means that its major focus was to explore the properties of numbers without concern for the real-world application of any of the results. Nonetheless, applications of number theory that are extremely important to the real world have resulted from research in this field. Cryptography, which is the transformation of information into a form that is unintelligible (and the reverse of this process) is commonly used in electronic transactions of all kinds to ensure privacy and security. Error checking codes, which are used in telephone communications, satellite data transfer, and compact discs, ensure that information remains intact. Both of these applications have foundations in number theory.

Fundamental Mathematical Concepts and Terms

Number theory is concerned with the properties of integers. Because of its concern with numbers, some people associate the terms arithmetic and higher arithmetic with number theory. Number theory is subdivided into a number of fields, the major ones being elementary number theory, analytic number theory, algebraic number theory, geometric number theory and Diophantine approximation. Several other fields of study within number theory include probabilistic number theory, combinatorial number theory, elliptic curves and modular forms, arithmetic geometry, number fields, and function fields.

Elementary number theory is one of the major subfields of number theory. The word elementary does not refer to the simplicity of the problems in this subfield, but rather to the fact that the problems studied do not use techniques from any other field of mathematics. Elementary number theory has a certain popular appeal because many of the problems are easily explained, even to people who are not mathematicians. However, finding solutions for these seemingly simple problems is often extremely complex and require great insight.

Some of the important problems involve prime numbers. Prime numbers are numbers greater than 1 that only have two divisors: 1 and the number itself. The prime numbers less than 10 are 2, 3, 5 and 7. As of 2005 the largest known prime number was $2^{25964951} - 1$, which has 2,816,230 digits. It is a special type of prime number

called a Mersenne prime. Prime number theory states that there are an infinite number of prime numbers; new ones are being found all the time.

Elementary number theory also investigates perfect numbers. Perfect numbers are numbers that are equal to the sum of all the integers that are its divisors. The number 6 is a perfect number. Its divisors are 1, 2 and 3 and the sum of these three numbers is 6. A second perfect number is 28. Its divisors are 1, 2, 4, 7 and 14, which sum to 28. Ancient Greeks discovered two more perfect numbers: 496 and 8,128. As of 2005, 42 perfect numbers were known and all of them were even. It is unknown if an odd perfect number exists, but if one does number theorists have shown that it will have at least seven different prime factors.

Questions of divisibility and prime factorization are also part of elementary number theory. Divisibility means that a number can be divided by another number without leaving a remainder. For example, both 5 and 6 are divisors of 30. Finding all the divisors that are prime numbers is prime factorization. The prime factors of 30 are 2, 3 and 5.

One of the important operators used in number theory is modulus. Modulus refers to dividing an integer by another integer and calculating the remainder. For example, 10 mod 2 = 0 because 10 is divided evenly by 2 and there is no remainder. In another example, 10 mod 3 = 1 because dividing 10 by 3 is 3 with a remainder of 1. Modulus is used in cryptography as described below.

The Euclidean algorithm is also part of elementary number theory. This algorithm is used to find the greatest common divisor of two integers. Euclid wrote it down in about 300 B.C., making it one of the oldest algorithms known. The greatest common divisor is the largest number that divides two integers without leaving a remainder. For example, the greatest common divisor of 42 and 147 is 21, although 3 and 7 are also common factors.

A second subfield of number theory is analytic number theory. This field involves calculus and complex analysis to understand the properties of integers. Many of these techniques depend on developing functions that describe the behavior of arithmetic phenomenon and then investigating the behavior of the function. This often makes use of the asymptotic nature of certain functions; functions that tend toward certain values called limits at extremely large (or small) values.

A number of statements in elementary number theory are easily described, but require extremely complicated techniques in analytic number theory to solve. For example, the Goldbach conjecture states that every even number greater than 5 is the sum of three primes. This conjecture has never been proven or disproven, but remains a source of much research in analytic number theory. The twin prime conjecture states that there are an infinite number of primes of the form p and $p + 2$. Although most mathematicians argue that this is true, it too has never been proven and remains an active area of research.

The subfield algebraic number theory concerns number that are algebraic numbers, which are numbers that are the solutions to polynomial expressions. All numbers that can be expressed as the ratio of two integers, also called rational numbers, are algebraic numbers. Some irrational numbers are also algebraic.

Some of the important areas of research in algebraic number theory are Galois theory, which studies how different solutions to polynomials are related to each other, and Abelian class field theory and local analysis, which investigate the properties of fields. In mathematics, fields are abstract structures in which all the elements can be subjected to addition, subtractions, multiplication and division (except by zero) and in which the distributive rule, the associative rule and the commutative rule all hold.

Geometric number theory and Diophantine approximation represent another field of study within number theory. Diophantus was an ancient Greek mathematician who lived in Alexandria, Egypt, probably in the third century A.D. He wrote a treatise called *Arithmetica* in which he described many problems concerning number theory. Diophantine equations are attributed to this great thinker and they are equations that have whole numbers as their solutions. Some of the most common Diophantine equations are whole number solutions to the Pythagorean theorem: $x^2 + y^2 = z^2$, which can also represent the length of the sides of a right triangle (a triangle that has one 90° angle). Some solutions include $3^2 + 4^2 = 5^2$ and $5^2 + 12^2 = 13^2$. In fact, Diophantus showed that there are an infinite number of whole number solutions to the Pythagorean equation. Other problems in geometric number theory incorporate the theory of elliptic curves, the theory of lattice points in convex bodies and the packing of spheres in different types of spaces.

Fermat's last theorem is one of the most famous statements in number theory. It claims that there are no solutions to the problem $x^n + y^n = z^n$ for any values of n greater than 2. In the margin of Diophantus's *Arithmetica*, the famous French mathematician Pierre de Fermat claimed "I have a truly marvelous demonstration of this proposition, which this margin is too narrow to contain." In 1665 he died without ever writing down the "marvelous demonstration." The statement was the

source of much fascination to mathematicians for more than three centuries. A cash reward was even offered to the person who could provide a proof of the statement. Most mathematician accept that an extremely complex proof using techniques in geometric number theory by mathematician Andrew Wiles finally proved Fermat's last theorem in 1994.

Real-life Applications

Number theory is a pure math discipline, which means it evolved without any attention to developing real-life applications. Nonetheless, number theory has proven to have real-life applications that affect almost everyone. As the Internet and other forms of electronic communication has become a larger part of daily life, the need to keep personal information private and to verify the identity of individuals becomes extremely important. Number theory provides techniques, which can be used to disguise information in order to ensure privacy and security. These techniques form the basis for the field of cryptography and they are used in a broad range of industries from retail stores to finance to government to healthcare. Every time a credit card is swiped, a bank transaction occurs, insurance agencies and hospitals send patient information to each other or the police use a driver's license to verify an identity, techniques from number theory are used to keep the information transferred secure.

While the goal of cryptography is making information harder to decipher, the goal of error correcting codes is to protect information from corruption. Error correcting codes are based in number theory and they are used in everything from the information beamed back to earth from Mars rovers to the compact disks that contain music.

CRYPTOGRAPHY

Cryptography is the set of techniques, usually mathematical, that are used to encrypt and decrypt information. Encryption means converting information from its understandable form to a form that is unintelligible. Most often, a set of mathematical steps called an algorithm is used for this purpose. A second algorithm is then performed to transform the unintelligible version of the message back to it original form. This is called decryption.

A simple example of encryption is the XOR algorithm. It can be used to transform binary codes. Binary codes are strings of 0s and 1s. All information in computers is eventually reduced to binary codes. Binary addition is slightly different from the addition that is commonly used with integers. It has four rules: $0 + 0 = 0$; $0 + 1 = 1$; $1 + 0 = 1$; and $1 + 1 = 0$.

Suppose that a binary message is 1010. A key for encrypting this message could be any string of four 0s and 1s; for example, 1101. Adding the original message to the key (bit by bit, with no carryover from the highest place—also known as the XOR function) results in an encrypted string. If someone were to intercept the encrypted string, they would not know what the original message was without the key. With symmetric keys, the same key is used for encryption and decryption.

There are several symmetric key systems in common use. The data encryption standard (DES) is one of the most popular, though it is not considered particularly secure because more than one person knows the key. The Diffie Hellman key agreement algorithm provides a higher degree of security because the parties involved in the exchange of information negotiate the key that they want to use as they exchange information. Because the key is developed as it is used, the chances that it will be intercepted by a third party decreases. In addition, the algorithm relies on the fact that the people involved in the negotiation will only have to do simple calculations to establish the key, but an eavesdropper would have to do very difficult calculations to steal it.

Encryption techniques that employ asymmetric keys, also called public keys, require that different keys be used for encryption and decryption. One of the most commonly used public key systems is the RSA Public-Key System. It is named for the last names of its developers, Ron Rivest, Adi Shamir and Leonard Adleman, who first developed the algorithm in the 1970s at MIT. These mathematicians build a company around the algorithm called RSA Data Security, headquartered in Redwood City, CA. RSA technology has been incorporated into a broad range of computer software including Microsoft Windows, Netscape Navigator, Intuit's Quicken, Lotus Notes, as well as operating systems for Apple, Sun and Novell computers. It is part of the Society of Worldwide Interbank Financial Telecommunications standards for financial transfers as well as standards used by the United States banking industry.

The RSA algorithm makes use of two important features of number theory: prime numbers and the modulus function. Its security depends on the fact that it is very hard to factor very large numbers. For example, the algorithm usually uses a modulus that is somewhere near 2^{800}; in order to discover the private key, an eavesdropper would need to find a way to factor the modulus.

Several types of factoring algorithms have been developed and they can be used to estimate the difficulty of

The RSA Public-Key Algorithm

The RSA algorithm is one of the most popular public key algorithms. It is probably best understood by example. Assume that a customer wants to make an electronic deposit of $3 using an automatic teller. The number 3 is the original message, *M*, which must be encoded for transfer and then decoded when the bank receives it. The automatic teller acts as the keymaker, generating numbers that act as keys for encryption and decryption.

In the first step of the RSA algorithm the keymaker generates two prime numbers: say $p1 = 11$ and $x2 = 2$. Next the product of the two numbers is calculated: $n = (p1)(p2) = (11)(2) = 22$. This number *n* is part of both the encryption key and decryption key. It is the modulus that is used later in the algorithm.

Next a number is calculated using Euler's totient function. This number is referred to as *t* and it is equal to $(p1-1)(p2-1) = (11-1)(2-1) = (10)(1) = 10$. The keymaker then selects a number, let's call it *e*, such that *e* is less than *t* and the greatest common divisor of *e* and *t* is 1. In this case the number chosen is 3, because it is less than 10 and the greatest common denominator of 3 and 10 is 1.

The next calculation requires finding a number *d* such that when the product of *e* and *d* is divided by *t* the remainder is 1. Another way to write this is $ed = 1 \bmod t$. In this case, *d* is 7 because $3 \times 7 = 21$ and $21/10 = 2$ with a remainder of 1.

The public key, used to encrypt the message is *e* and *n*, in this example 3 and 22. The keymaker may make this key known to everyone. The private key is *d* and *n* or 7 and 22, and only the keymaker knows it.

The keymaker now transforms the message by raising the message *M* to the power *e*, dividing by *n*, and calculating the remainder. This calculation can also be written $M^e \bmod n$. In this example $M^e = 3^3 = 27$. The number 27 is divided by $n = 22$ and the remainder is 5. The encrypted message, *E*, is 5.

The automatic teller then sends the encoded message ($E = 5$) to the bank, along with the private key, which in this example is *d* and *n* or 7 and 22.

The encrypted message and the private key is received by the bank, they decrypt it using the calculation, $E^d \bmod n$. In this example, $E^d = 5^7 = 78{,}125$. The number 78,125 is divided by $n = 22$ and the remainder is 3, which was the original message.

Although very small numbers were chosen for *p*1 and *p*2 in this example, in practice they are usually on the order of 2^{400}, which makes *n* extremely large, somewhere near 2^{800}. The difficulty in factoring such big numbers is crucial to the security of the algorithm. If the factors of *n* were easy to find, then discovering the private key would not be that hard to do.

factoring very large numbers. According to one of these algorithms, in order to factor a value number that is close to 2^{800} would require about 2^{77} steps. In 2005, the average computer could do about 100 million instructions per second. This corresponds to 100,000,000 instructions/second $\times$ 60 seconds/minute $\times$ 60 minutes/hour $\times$ 24 hours/day $\times$ 365 days/year $= 3 \times 10^{15}$ instructions/year, which can also be written approximately as 2^{51} instructions per year. As a result it would take about $2^{77} / 2^{51} = 2^{(77-51)} = 2^{26}$ or roughly 70 million years to factor the modulus.

In addition to RSA, other groups of public key algorithms have been developed. One is called ElGamal and it relies on similar mathematics as the RSA algorithm. ElGamal can also be used to verify that information sent has not been compromised during transmission. It does this by means of a digital signature and special mathematical functions called Hash functions. Digital signal algorithm (DSA) can also be used for digital signatures. Another public key algorithm relies on functions called elliptic curves, which are studied in number theory and have become increasingly popular for use with cryptography.

ERROR CORRECTING CODES

As binary information (information coded as strings of 0s and 1s) is transmitted, errors can occur in the string, which make the information unintelligible. Error correcting codes are algorithms that ensure that information is transmitted error-free and many of these algorithms depend on results from number theory.

Claude Shannon and Richard Hamming working at Bell Laboraties in the late 1940s developed a method of repeating strings to ensure that the information sent was received. They worked out theories, which optimized the number of repetitions necessary to ensure that the information received was correct. Another researcher,

Key Terms

Algorithm: A set of mathematical steps used as a group to solve a problem.

Binary code: A string of zeros and ones used to represent most information in computers.

Decryption: The process of using a mathematical algorithm to return an encrypted message to its original form.

Divisibility: The ability to divide a number by another number without leaving a remainder.

Encryption: Using a mathematical algorithm to code a message or make it unintelligible.

Greatest common divisor: The largest number that is a divisor of two numbers.

Integer: The positive and negative whole numbers.. –4, –3, –2, –1, 0, 1, 2, . . . The name "integer" comes directly from the Latin word for "whole." The set of integers can be generated from the set of natural numbers by adding zero and the negatives of the natural numbers. To do this, one defines zero to be a number which, added to any number, equals the same number.

Key: A number or set of numbers used for encryption or decryption of a message.

Modulus: An operator that divides a number by another number and returns the remainder.

Perfect number: A number that is equal to the sum of its divisors.

Prime factorization: The process of finding all the divisors of a number that are prime numbers.

Prime number: Any number greater than 1 that can only be divided by 1 and itself.

Public key system: A cryptographic algorithm that uses one key for encryption and a second key for decryption.

Symmetric key system: A cryptographic algorithm that uses the same key for encryption and decryption.

John Leech, also developed theories related to error correcting codes. His work included some abstract mathematical in number theory such as groups and lattices.

Error correcting codes were put to immediate use at NASA, where satellites were equipped with powerful error checking codes. A typical algorithm of this type is capable of correcting seven errors in every 32 bits sent back to earth. The redundancy in the data sent is immense; in every 32 bits only 6 are data. The rest are for error checking.

When they were initially developed, compact discs were highly sensitive to scratching and cracking. But by incorporating two redundant codes that are interleaved, CD players can recover up to 4,000 consecutive errors. Additional error checking algorithms are built into CD players to further correct problems with the signal.

Where to Learn More

Books

Gardner, Martin. *The Colossal Book of Mathematics*. New York: W.W. Norton & Company, 2001.

Gullberg, Jan *Mathematics: From the Birth of Numbers*. New York: W.W. Norton & Company, 1997.

Hoffman, Paul. *The Man Who Loved Only Numbers*. New York: Hyperion, 1998.

Paulos, John Allen. *Beyond Numberacy: Ruminations of a Numbers Man*. New York: Alfred A. Knopf, 1991.

Periodicals

Schroeder, Manfred R. "Number theory and the real world," Math. *Intelligencer 7* (1985), no. 4, 18–26.

Web sites

Department of Mathematics University of Illinois at Champaign Urbana. "Guide to Graduate Study in Number Theory." January 23, 2002. <http://www.math.uiuc.edu/ResearchAreas/numbertheory/guide.html> (April 27, 2005).

Grabbe, J. Orlin. "Cryptography and Number Theory for Digital Cash" October 10, 1997 <http://www.aci.net/kalliste/cryptnum.htm;> (May 2, 2005).

The Mathematical Atlas. "Number Theory." January 2, 2004. <http://www.math.niu.edu/~rusin/known-math/index/11-XX.html> (April 27, 2005).

Pinch, Richard. "Coding theory: the first 50 years" Plus Magazine. September 1997. <http://pass.maths.org.uk/issue3/codes/> (May 4, 2005).

The Prime Pages "The Largest Known Primes." <http://primes.utm.edu/largest.html> (May 9, 2005).

RSA Security. "Crypto FAQ" 2004. <http://www.rsasecurity.com/rsalabs/node.asp?id=2152> (May 4, 2005).

Weisstein, Eric W., et al. "Number Theory." MathWorld—A Wolfram Web Resource. <http://mathworld.wolfram.com/NumberTheory.html> (April 24, 2005).

Odds

Overview

Probability is a form of statistics used to predict how often specific events will occur, and is used in fields as varied as meteorology, criminal justice, and insurance underwriting. When probabilities are calculated, they are frequently expressed in terms of odds. Odds provide a simple, shorthand language for communicating probabilities, regardless of the specific situation being assessed. Odds can be expressed using differing terminology and notations, but the basic principles remain constant, regardless of the application.

Fundamental Mathematical Concepts and Terms

Several systems of terminology can be used to express the odds of a particular event occurring. Consider the case of a standard deck of 52 playing cards, which consists of four suits of 13 cards apiece. A dealer takes this deck, shuffles it thoroughly, then draws a single card; what are the odds that he will draw the single Ace of Hearts? The odds of drawing this particular card out of the fifty-two in the deck are denoted 1:52. This probability can also be described as one chance in 52 of successfully drawing the desired card, or expressed as a fraction: 1/52.

Odds are also expressed in reverse, giving the odds against an event happening. In the previous example, the odds of drawing the desired card from the deck could also be expressed as 51:1, meaning that of the 52 possible outcomes, 51 would be undesirable while only 1 would be the hoped-for outcome. In some cases, the odds for and the odds against are used interchangeably: odds of 1 in a million and odds of a million to 1 are both used to describe extremely unlikely events, and are almost identical mathematically. However, this same relationship does not hold true for smaller values, with odds of 1 in 3 (33%) being significantly better than odds of 3 to 1 (25%).

A Brief History of Discovery and Development

Because odds are simply the language of probability, the history of odds runs parallel with the history of probability, and is discussed extensively in the entry on that subject. However, as the language of odds has been applied to an expanding array of applications, a unique vocabulary has developed around the use of odds. Unfavorable odds, such as odds of one in a million, have come

to be called long odds, meaning the event they describe is highly improbable. Long odds are also sometimes referred to as a long shot; slow race horses and unknown political candidates are often described as long shots, suggesting that their odds of winning are remarkably small. Odds are sometimes expressed in terms of a percentage, or on a base of 100. A salesman who claims to be 90% sure he can deliver his product on time is offering odds of 9 in 10 that he will succeed. A wildcatter drilling a new oil well might give odds of 60/40 that the new well will be a gusher, placing the odds at 6:4, which can be reduced to 3:2 and then reduced further to 1.5 to 1 that he will succeed. A stock analyst who gives a stock a 50:50 chance of rising is giving it a 1 in 2 chance, equal to the chance of flipping heads on a single toss of a coin.

In a few cases, odds are used to imply that an event is absolutely certain to occur; theoretically these odds would 1:1, or 100%, though the certainty of any future event is always less than 100%. However in these cases, an event is often referred to as a lock, or a sure thing, suggesting that it will certainly occur. However the history of gambling and athletics is replete with sure things which failed to materialize, suggesting that the sure thing and the lock are more a result of wishful thinking than of rigorous statistical analysis.

Real-life Applications

SPORTS AND ENTERTAINMENT ODDS

Poker is a popular card game in which odds are used to develop strategy, and successful poker players often possess an innate sense of the odds associated with certain hands. In the course of a typical poker game, players are often forced to make quick decisions on whether a hand is winnable and should be played, or is unwinnable and should be folded. For example, a player holding a 5, 6, 7, 9, 9 must decide whether to keep the 9s and hope to be dealt a third 9, or to discard one of the nines in the hope of drawing an 8 to complete a straight. Using a basic understanding of probability and the rules of the game, an experienced player will probably keep the 9's, knowing that the odds of ending up with a winning hand are significantly better using this strategy.

Gambling in any form has been a popular pastime for most of recorded history. During the past century, gambling, or gaming, as it is sometimes called, grown from a casual pastime into a multi-billion dollar industry. As the gambling industry has grown, casino owners have increasingly turned to fields such as psychology and marketing in order to increase their earnings. The very existence of these extravagant entertainment centers, some costing more than $1 billion to build, simply confirms the efficiency with which casino owners separate players from their money.

Modern casinos are scientifically designed to lure players in and keep them playing as long as possible. A typical casino contains a variety of games, offering a wide array of playing styles and varying odds of winning though one fact remains: in every form of casino gambling, the odds of the game favor the casino, or as it's known in the industry, the house. In most cases, the tilt in favor of the house is slight, allowing some players to beat the house over the short-run and leading to impressive tales of huge jackpots. But the ultimate result is the same as in any other activity governed by the laws of probability: over the long-run, the house will always win.

The house edge, or how strongly the odds of a particular game favor the casino, vary from game to game. The game of roulette, in which a ball is dropped onto a spinning number wheel, offers a house edge of 5.6%, meaning that in practice, for each $100 wagered, a player will lose an average of $5.60. If a roulette player spends two hours playing, betting $25 per spin and averaging 30 spins per hour, the casino will expect to make about $75.00 in that time, and the customer will have paid roughly $37.50 an hour for the privilege of watching a small marble drop onto a shiny spinning wheel. Of course two other outcomes are also possible. A player could actually win several times in a row and walk away with his winnings, taking the house for a loss; this possibility is what keeps die-hard gamblers coming back for more action. The other possibility is that the player hits a run of tough luck and loses his entire stake sometime during the session. In this case, the casino's edge has simply been felt earlier than expected, and the player goes home empty-handed.

Roulette offers some of the lowest odds of any casino table game, meaning the house edge is larger in this game than in most others. Blackjack, a card game in which a player tries to collect cards totaling 21, offers a theoretical house edge of only 0.80%, though in practice few players play the game with such computer-like precision, making the actual house edge higher. Assuming a gambler can follow the optimal betting strategy without error, he should be able to wager $100 during his session and lose only 80 cents to the casino. Similar odds accompany the game of craps, in which dice are rolled and games are won and lost based on the outcome of the roll.

Some of the worst odds in the casino are offered by one of the most popular games, the slot machine. Aptly nicknamed the "one armed bandit," these flashing, beeping machines involve no skill whatsoever, requiring

players only to insert a coin and pull a lever or push a button. Based on the outcome of a set of spinning wheels, prizes are paid according to a table on the front of the machine. The house edge for slot machines is difficult to calculate, because machines can be programmed to return a higher or lower amount of player money. As a general rule, slot machine payouts vary depending on the amount required to play, with higher play values receiving better odds. The typical house edge for a nickel slot machine would be somewhere near 8%, meaning that for each $100 bet, a player would typically lose about $8.00. Odds are much better on higher-value slot machines, meaning that a player willing to spend $5.00 per play will face a less severe house edge. Unfortunately, he will also burn through his funds much more quickly. Slot machines offer high efficiencies to casino operators. By combining low operating costs, a high house edge, and the potential for players to bet several times per minute, one armed bandits are among the casino's best moneymakers, probably explaining why most Las Vegas casino entrances are lined with a sizeable collection of the shiny machines.

Why do people gamble? Few other activities offer a guaranteed chance to lose money, yet gambling today is more widespread in both the physical and the virtual world than ever before. For some players, gambling is simply a form of entertainment. These players typically allot a set amount to spend on an outing, wager and enjoy the experience and the excitement, then leave, having paid relatively little for their entertainment. For other players, gambling is perceived as a chance to improve their lot in life by offering a fast route to large amounts of cash.

Some percentage of gamblers behave irresponsibly, wagering far more than they can afford to lose and creating serious problems for themselves and their families. Compulsive gamblers are similar to compulsive drinkers in that they are unable to moderate their behavior; in some cases, compulsive gamblers spend entire paychecks or close out bank accounts attempting to recoup previous gambling losses. Gamblers Anonymous, an organization created to help compulsive gamblers recover, provides a list of questions to help gamblers determine whether they have a problem. Questions such as, "Did you ever gamble to get money to pay debts," "Have you ever sold anything to finance gambling," "Did you ever gamble down to your last dollar," and "Did you ever have an urge to celebrate good fortune with a few hours of gambling?" are intended to help gamblers assess their situation. According to the National Council on Problem Gambling, in 2005 between 3,000,000 and 12,000,000 Americans had gambling problems of varying degrees.

The game of roulette offers a house edge of 5.6%, meaning that in practice, for each $100 wagered, a player will lose an average of $5.60. ROYALTY-FREE/CORBIS.

ODDS IN EVERYDAY LIFE

While the question "paper or plastic?" is probably the most common dilemma faced by shoppers, one shopper's quandry (perplexing question) is as old as the supermarket: how can a shopper tell which check-out line will move fastest? Obviously if one line has fewer shoppers in it, that is probably the line to choose, though smart shoppers also know that if the light above that line's cashier is flashing it's probably a danger sign. But what if all the lines have the same number of customers waiting? What chance is there of choosing the fastest line? All things being equal, a shopper's chance of choosing the fastest line is actually pretty small, a statistical reality born out by many shoppers' frustrating experiences.

Consider this quandary as a probability question. Assume that the shopper has no information about which checkers currently working are faster or slower, meaning that in this case, his decision comes down a random selection. Once he has made his choice, in this case choosing lane three, he will be forced to stand and watch the other lanes to learn whether or not he chose wisely. From lane three he is able to see lanes one and two on his left and lanes four and five on his right, and chances are good that he will soon find out he did not choose the fastest lane. A simple odds calculation explains why.

The total group of possible outcomes consists of the five cashiers from which the shopper can choose, while the outcome of interest is the particular lane the shopper ultimately selects. All other factors being equal, the odds of choosing the fastest line are 1 in 5; put more pessimistically, the odds of choosing incorrectly are 4 in 5, meaning that most days, most shoppers will watch at least one other line move faster than the one they have chosen. Given these poor odds, one might instead opt for a new

development, self check-out, which in most cases is markedly slower than being checked out by a professional, but which eliminates the annoying wait in line, as well as providing a reasonable distraction from the process.

Human beings have an innate fascination with events or objects which seem to defy the laws of probability. Chinese basketball star Yao Ming fascinates most Westerners, not just because he is a basketball star or even because he is an unlikely 7'6" tall. Most Americans are stunned by Yao's towering height because he beat the odds by growing so tall in a nation where the average man is around three inches shorter than the average American man. While Yao seems like the ultimate example of a long shot, two factors make his height far more understandable. First, his parents are exceptionally tall, even by U.S. standards, with his father standing 6'7" tall and his mother measuring 6'3". Second, regardless of the average national height, China is more likely to produce another giant player simply by virtue of its enormous population: with more than 1.3 billion residents, mainland China is more than four times as populace as the U.S., dramatically boosting its chances of producing another seven-footer or two. In addition, both the U.S. and China are demographically diverse, and there are areas of China where the average height exceeds the average world and U.S. height for males.

ODDS IN STATE LOTTERIES

Many states and countries now generate revenue by operating lotteries. In a typical lottery, players are encouraged to buy a ticket with a set of numbers on it, such as six numbers between 1 and 45. Tickets are sold for a set period of time, then a drawing is held in which 6 numbers are randomly selected. Matching some of the selected numbers is rewarded with a cash prize determined by the number of numbers guessed correctly. In a typical lottery, the largest prize, the jackpot, is won by correctly guessing all the numbers selected, with the prize normally being more than one million dollars. Because most lotteries are run by government agencies, they are required to publicize details of the games, such as the odds of winning at each level and the actual use of the lottery's earnings.

State lottery managers must make several decisions in order to maximize the number of players and, by extension, the total number of dollars earned for the state; for this reason, lottery rules change frequently in order to keep players interested. One state lottery in 2005 used the following formula: two sets of numbers are used, each running from one to 44. From the first set of numbers, five values are randomly selected, and from the second set of numbers, a single bonus number is chosen. A player lucky enough to match all five of the initial numbers has managed to beat odds of 1.1 million to one and will collect a sizeable prize. But a player who manages to combine this feat with a correct pick of the bonus number wins the top prize, frequently in the tens of millions of dollars. Jackpots often go unclaimed for several weeks, since the odds of picking all six values correctly are 1 in 47 million. A player's odds of winning any prize (prizes start at $3.00) in a single play are 1 in 57.

Where does lottery money go? Lottery proponents are quick to point out that the net proceeds of a lottery are typically spent on education and other popular projects. However, the actual education income from lottery tickets is typically less than one-third of the money spent playing the game. The Texas State Lottery in 2004 published a breakdown of how its income was spent. For each dollar wagered, the program returned fifty-two cents to players in the form of prizes, making lotteries among the worst bargains in gambling when compared with almost any casino game. Seven cents of each lottery dollar also went to administrative costs associated with running the lottery itself, including salaries for administrators and advertising costs, while another five cents was paid to retailers in return for their work selling the tickets.

Once all these costs are removed, this particular lottery program contributes the remaining thirty cents of each dollar to the state's education agency for use in local school programs. Is this level of return high or low? The answer depends on whether the lottery is analyzed as a business or as a non-profit fund-raising agency. For most business CEOs, managing to pass 30% of their gross revenues along to their owners would make them among the most successful and admired business leaders in the hemisphere. But a non-profit fund-raising organization which consumes 70% of its revenues paying administrative and other costs before passing less than one-third of contributions along to its beneficiary is generally considered either unethical or grossly incompetent. Because state lotteries fail to cleanly fit either model, the appropriateness of this 30% pass-through rate remains controversial.

Lotteries have risen from relative obscurity in America during the early twentieth century to a point where most states operate the programs. Several facts have contributed to this rise in popularity. One trend which helped lotteries flourish was a general resistance to additional taxes, beginning with the Reagan presidency in the 1980s and continuing through the turn of the century. In an atmosphere where even proposing a tax hike could be politically fatal, lotteries provided a sizeable revenue boost without the political costs of a tax hike. A second

argument has gained steam as lotteries have spread to cover most of the country. Economic studies have looked at out-of-state revenue garnered by lotteries, concluding that in many cases, residents of non-lottery states drive across the border to buy tickets when jackpots grow large. In response, some non-lottery states eventually conclude that initiating their own lottery is the lesser of two evils when compared with continuing to watch residents take their dollars to neighboring states.

A final argument in favor of starting lotteries is the voluntary nature of participation. For many voters, the choice between a hike in property taxes, which impacts most citizens, and a state lottery which produces the same amount of revenue but is purely voluntary, may seem straightforward. However, given the lottery's common nickname of "the stupid tax," (a perspective on the fact that the "tax" is regressive in that it is generated even if voluntarily) from people with lower levels of education and income as opposed to the general population.

While the long-term impact and success of state lotteries remains to be proven, various challenges have already faced lottery administrators. For example, in one case a larger than expected number of players of one particular game won jackpots during the first quarter of the year. Because lottery profits are tied directly to the number of tickets sold, this statistical fluke did not endanger the lottery's solvency. However, because of the psychological appeal of larger prizes, this run of smaller jackpot winners did substantially reduce the number of ticket sold, slashing income during the first months of the year. By mid-year, directors were evaluating a change in the game rules to reduce the number (and increase the size) of jackpots won.

OTHER APPLICATIONS OF ODDS

Large numbers are used for a variety of business purposes, including security. A typical credit card uses a sixteen digit account number. For a thief trying to make an online purchase, how hard would it be to simply guess random numbers until he chose one that was valid? The total number of credit card account numbers possible using all sixteen digits is 10^{16}, meaning that the odds of guessing a particular number on a single try are 1 in 10,000,000,000,000,000. Since most credit cards start with the same few sets of four digits, those digits are not available for creating account numbers, reducing the number of possible account numbers to 1 in 10^{12}, or 1 in 1,000,000,000,000. If United States consumers held one billion credit cards, the odds of guessing a correct number would improve dramatically, to roughly 1 in 1,000. Using modern software, a thief could easily try 1,000 numbers in order to find one that would work; for this reason, most credit card companies now also require a three to four digit code from the back of card, as well as complete personal information including the billing address, making the guessing tactic relatively useless. In response, twenty-first century thieves are far more likely to focus on hacking into massive credit databases where they can steal millions of valid card numbers and billing addresses at one time, rather than spending time guessing card numbers.

While providing a set of odds lends an air of credibility to a claim, odds are sometimes assigned to an event based on little evidence. Surgeons and other care-givers are frequently asked to give worried family members the odds of a patient's recovery from a serious illness. How can a concerned doctor provide these odds? An experienced surgeon can probably scan his memory for similar cases, or refer to medical reference volumes that estimate the likelihood of recovery. Unfortunately, given the wide variations in actual patient conditions, these methods are subject to wild swings in accuracy, and in many cases, a doctor is probably forced to provide an estimate based on little more than his instincts.

Ironically, in the case of a doctor giving odds for a patient's survival, some incentive exists for the doctor to actually overstate the danger and give lower odds than he might otherwise. For example, imagine that a doctor assigns a serious case 1 chance in 10 of surviving; if the patient dies, the concerned family and friends will likely be unsurprised, since the odds provided were not favorable. On the other hand, if this long shot patient manages to pull through, the family will be elated, potentially concluding that doctor is a medical genius. In this case, with the doctor providing unfavorable odds, he is ultimately perceived more favorably whether the patient lives or dies.

Conversely, consider the situation in which a doctor provides an overly optimistic assessments. If a doctor gives the patient survival odds of 9 in 10, the patient's recovery will occur only as an expected event. However, if the patient takes a turn for the worse and ultimately dies, the stunned family will have been emotionally unprepared, and may in fact blame the doctor or file lawsuit based on their expectations rather than the merits of the case. Subconsciously, the doctor who consistently gives his patients good odds only to watch them die may begin to question his own performance. While most doctors undoubtedly try to give accurate assessments of a patient's prognosis, little incentive exists for a doctor to give optimistic assessments. Subconsciously, this may lead doctors to paint a grim picture while hoping for a positive outcome.

Key Terms

80/20 rule: A general statement summing up the tendency for a few items to consume a disproportionate share of resources, such as cases in which 20% of a store's customers lodge 80% of the total complaints.

Long odds: Poor odds, or odds which suggest an event is highly unlikely to occur.

Odds: A shorthand method for expressing probabilities of particular events. The probability of one particular event occurring out of six possible events would be 1 in 6, also expressed as 1:6 or in fractional form as 1/6.

Probability: The likelihood that a particular event will occur within a specified period of time. A branch of mathematics used to predict future events.

Doctors are not alone in struggling to understand the odds of common events. Despite numerous news stories on the topic, many people still believe that flying is a more dangerous form of travel than driving. In truth, the odds of dying while driving to the airport are frequently higher than the odds of dying in the ensuing plane trip. People's irrational fear of flying is perhaps because hundreds of automobile accidents occur each day without notice, while every passenger plane crash receives extensive media coverage.

A similar situation exists in the energy industry. Many people believe that nuclear energy is the most dangerous form of power generation, and that the odds of being killed or injured in a nuclear power accident are far higher than the odds of being injured by processes related to coal or other low-tech energy sources. Ironically, far more people have died as a result of working in coal mines or in the coal industry than from the U.S.'s single nuclear power accident at Three Mile Island. Each year, a person's odds of being killed by a simple electric shock or by falling down in the bathtub are far higher than her odds of being exposed to dangerous levels of radiation.

Because most people have such a hard time assessing the odds of an event occurring, some companies market products designed to allay people's fears of these events occurring. Insurance companies, well-aware of both the slim odds of an airplane crash and the public's irrational fear of flying, have long offered flight insurance policies at airports, comfortable that these policies will provide both peace of mind to their buyers and steady income to their sellers.

In some situations, people take elaborate precautions to protect against hazards with relatively low odds, while ignoring other events with much higher odds of occurring. Most responsible drivers recognize the hazards of driving while intoxicated and avoid this behavior because of its potential for disaster. However, the number of drivers who talk on their wireless phones while driving remains high, despite advertising campaigns informing drivers that the chance of being in an accident are roughly the same in both situations. While the odds may be the same in both cases, most chatting drivers apparently do not recognize the hazard they are creating.

Because the language of odds provides a shorthand way of discussing possible outcomes, informal odds are frequently assigned to events as part of a discussion. One example of this informal use of odds is the well-known 80/20 rule. While not a strict mathematical set of odds, this rule is commonly used to explain a variety of situations in which some small portion of the whole has a larger than expected influence on the outcome. For example, managers sometimes use this rule in describing employees who require exceptionally large amounts of time, implying that the trouble-prone 20% of their employees are responsible for 80% of the manager's total problems. Teachers sometimes use this same rule to describe certain students who require constant assistance. Fund-raisers who spend their careers soliciting donors for causes such as education, disease research, or religious work, are well aware that the 80/20 rule accurately approximates the distribution of their donors, with the most generous 20% contributing 80% of the total funds, while the other 80% collectively give only 20%. Inventory managers frequently observe a similar pattern, with a relatively small number of products making up a significant majority of total order volume received.

The 80/20 principle can guide decision-making in a variety of scenarios. A fund-raiser who recognizes the pivotal role played by his larger contributors will go to extraordinary lengths to remain in favor with these donors, since he knows that a large donor is both far more painful to lose and far more difficult to replace. In optimizing inventory management, a warehouse supervisor can observe that 80% of orders received will include

at least one of the top 20% of products, meaning he will keep a sizeable supply of these items on hand for immediate shipment. On the other side of this equation, this same manager knows that he can safely inventory smaller quantities of the less popular items, since the odds of any single one being ordered on a given day is quite small.

In a few cases, people can improve the odds of favorable outcomes. A male college applicant might consider the ratio of men to women at a particular campus, assuming that a larger number of single women on a campus would raise his odds of finding an appropriate date or mate. While the ratio of men to women is roughly 1:1 in the entire world, specific locations feature some surprisingly large variations in this mix.

Potential Applications

Because odds are simply the language of probability, future developments in the use of probability will be reflected in future uses of odds. With the growing popularity of poker as both a participant and a spectator sport, odds may become a more common topic of discussion, and fans of the game may commit the odds of completing particular hands to memory. In addition, recent advances in computer technology will likely lead to more accurate assessments of odds for numerous events, such as predictions of the future state of the U.S. and world economies.

Where to Learn More

Books

Epstein, Richard A. *The Theory of Gambling and Statistical Logic.* New York: Academic Press, 1977.

Glassner, Barry. *The Culture of Fear; Why Americans are Afraid of the Wrong Things.* New York: Basic Books, 1999.

Orkin, Mike. *What Are the Odds? Chance in Everyday Life.* New York: W.H. Freeman and Co., 2000.

Web sites

BBC News World Edition. "Clumsy Horse Beats all the Odds." <http://news.bbc.co.uk/2/hi/uk_news/wales/4377413.stm/> (March 30, 2005).

CNN Law Center. "Chinese Height Discrimination Case." <http://www.cnn.com/2004/LAW/05/31/dorf.height.discrimination/> (March 27, 2005).

National Council on Problem Gambling. "What is the Biggest Problem Facing the Gambling Industry?" < http://www.ncpgambling.org/media/pdf/g2e_flyer.pdf> (March 29, 2005).

Percentages

Overview

A percentage is a fraction with a denominator of 100. A percentage may be expressed using the term itself, such as 25 percent, or using the % symbol, as in 25%.

The calculation of various kinds of rates by way of percentages is the backbone of a wide range of mathematical applications, including taxes, restaurant tips, bank interest, academic grades, population growth, and sports statistics.

Fundamental Mathematical Concepts and Terms

Percentages are the natural mathematical extension of three other familiar concepts: fractions, ratios, and proportions. A fraction is a number expressed as one whole number divided by another; for example, one half is expressed as ½. A ratio is the relationship between two similar magnitudes. For example, as of 2005, the relationship between the population of Canada, estimated at 31 million people, and that of the United States, measured at approximately 310 million people, is a ratio of 1 to 10.

A proportion is a pair of ratios expressed as a mathematical equation. For example, if in a city of 100,000 residents, 1,000 people had red hair, the proportion of the population with red hair will be expressed as 1,000/100,000, or 1/100. The equation 1,000/100,000 = 1/100 is a proportion.

All percentages are an expression of a relationship based on 100. Every fraction, ratio, and proportion may be expressed as a percentage. Percentages may also be expressed where decimals are required, as in the figure 66.92%.

An important application of the concepts concerning percentages is that of percentiles. A percentile, which is one of the 99 points at which a range of data is divided to make 100 groups of equal size, is an important tool used in a vast number of statistical areas. For example, students in a class or across a larger population are given percentile rankings on a national test. The determination of the percentile ranking is a way of measuring relative standing to every other person in the class or larger group.

DEFINITIONS AND BASIC APPLICATIONS

A percentage is a fraction with a denominator of 100. A percentage may be expressed in any of the following ways: 38 percent, 38%, 38/100, or 0.38 as a decimal notation.

- To convert a decimal to a percentage, move the decimal point two places to the right and tack on a % sign. For example, the decimal 0.09 equals 9%.
- To convert a percentage to a fraction, remember that x% will always mean x/100; for example, 40% = 40/100. The simplest form of this fraction is 4/10, or 2/5.
- To convert a fraction to a percentage, find the decimal equivalent of the fraction and convert the decimal to a percentage as described above. For example, 3/4 = .75 = 75%.
- Finding a percentage of a quantity is common. For example, 15% of a shipment of 350 books is stated to be damaged by water. To find the actual number of damaged books in the shipment, proceed as follows: the word "of," as used here, means multiply, or, (15/100) × 350 = 52.5.

RATIOS, PROPORTIONS, AND PERCENTAGES

To properly understand the many ways that percentages can be applied in modern life, it is important to understand the relationship between ratios, proportions, and percentages. These terms are commonly applied, and each has a separate meaning and a distinct mathematical purpose.

A ratio is defined as the expression of the relative values of numbers or quantities, using one of three forms:

- use of the word "to," as in "a ratio of 8 to 5"
- use of a colon, as in 8:5
- use of a fraction, as in 8/5

A ratio may also be expressed where different quantities are related. For example, the relationship of 20 minutes to one hour is the same relationship as 20 minutes to 60 minutes, or a ratio of 20:60, or 2:6, and ultimately, a ratio of 1:3.

Other examples of ratio conversion include 5 tons to 500 pounds, 10,000 pounds to 500 pounds, 10,000/500, and a ratio of 20:1.

Ratios are a common form of expression in certain forms of sports wagering and games of chance. When a certain horse is favored to win at a racetrack, the probability of that horse winning its race, referred to as the odds of winning, is expressed in ratio form, for example, 3 to 2. In this context, the ratio means that for every two dollars agreed, the bettor will win three dollars if the horse wins.

The calculation of odds finds itself in other aspects of daily living. If in a particular place, over the course of an average year, 35 young drivers (under the age of 21) out of a sample of 100 young drivers were involved in motor vehicle accidents, and 10 older drivers (over the age of 50) were involved in accidents, what are the odds of a young driver being involved in accident versus those of an older driver? The odds are calculated as follows: 35/65 × 10/90 = 4.85. Therefore, the odds of the young driver being involved in an accident might be said to be almost 5 (rounding up the 4.85 figure).

Proportions result when two ratios are set equal to one another. For example, 6:9 = 12:18: a/b = c/d.

A Brief History of Discovery and Development

The term percent is derived from two Latin words: *per*, meaning by, and *cent*, meaning one hundred. The use of multiples of 10 as the basis for arithmetic, the forerunner to the modern decimal system, first gained acceptance with the Pythagorean school of mathematicians based in Greece in approximately 400 B.C.

However, the percentage is a relative latecomer as mathematical developments are gauged. The decimal had been developed as an effective way to easily distinguish between fractions with different denominators (for example, on first observation, the fractions 4/13 and 5/17 have similar values, but the corresponding decimal conversions for each, 0.307 and 0.294, are clearly different values). The decimal point became standard throughout the European scientific community in the early 1600s.

The introduction of the decimal fraction was one of the great advances of mathematics. This occurred because the decimal simplified numerical calculations, thus engineers, surveyors, and scientists could express their work to any desired degree of accuracy. The decimal fraction eliminated the potential for errors when fractions were compared with one another or converted in the course of measuring or other mathematical calculations.

The percentile concept was first developed in 1885 by English physician and mathematician Sir Francis Galton (1822–1911). His motto, "Whenever you can, count" is as appropriate today as when Galton coined the expression, given the role of the percentile in the modern world's obsession with measuring and ranking an infinite range of activities, from business to government to sport.

The percentage is now used as both a general descriptive term (in phrases such as "play the percentages," "there is no percentage in that"), as well as a mathematical tool of comparison and analysis.

The understanding of the various ways that percentage calculations may be used is crucial to the successful

navigation of commercial, academic, and social worlds. Because society is now so accustomed to percentages being advanced in support of a particular viewpoint or concept, percentages can sometimes convey a superficial or misleading sense of certainty about a topic. Broad statements made in the media by business leaders, government officials, and others speaking on public issues often incorporate the expression of percentages. An example, "The economy will grow by 2% this year," has the ring of authority because a specific figure, 2%, is stated. However, an understanding as to how a particular percentage figure was arrived at is more important than the figure itself.

Similarly, an NBA basketball player may take pride in making 65% of his shots in the course of a playing season. If he only takes five shots per game, when his team is regularly scoring 100 points or more per game, the superficial impression and the impact of the high shooting percentage is much less, and the 65% figure is deceptive.

Real-life Applications

IMPORTANT PERCENTAGE APPLICATIONS

Percentages are calculated in a multitude of real-life situations. The understanding and proper applications of various percentage calculations are critical to daily living. The most relevant of these applications are set out below:

- The calculation of any type of rate: bank interest rate, a student loan rate, tax rate, mortgage rate.
- Education: determination of student grades. The ranking of students will often be determined by their grades, usually expressed as a percentage, as well as determined by the calculation of a related application to that of the percentage, the determination of a percentile.
- Science: in fields such as chemistry, pharmacology, or medicine, it is essential to be able to calculate the concentration of a particular substance in a mixture or solution.
- Food industry: percentages are used to determine the relative amount of the contents of food and beverage products, including the amount of certain fats, the amount of alcohol by volume in liquors, and the amount of a recommended vitamin or mineral.
- Retail sales: pricing increases and sales discounts are almost always expressed as a percentage. It would be difficult for businesses and customers alike if a price reduction was expressed as 2/7 of every dollar off, as opposed to a percentage figure.
- Social studies: any analysis of population growth, income, spending, inflation, or unemployment is expressed as a percentage.
- Meteorology: weather forecasts express the possibility of certain changes in the weather as a percentage; for example, a 20% possibility of precipitation.
- Sports: percentages are used to make comparisons in all types of competition. The shooting percentage in basketball, a quarterback's pass completion percentage in football, or baseball's batting average have become essential to the manner in which these sports are understood.
- Business: a company's current performance, the prospects for future growth, and measures of profitability and returns on investment will all be measured by percentage applications.
- Government and public service: trends in government spending, the increases or decreases in all aspects of the size, nature, and extent of public service, and future projections of every kind.

EXAMPLES OF COMMON PERCENTAGE APPLICATIONS

The 1% method This method of calculation is often useful for quickly determining small percentages. Determine 1% of the given number, and then compute the value of the desired percent.

To calculate 3% of 1,800: if 1% of 1,800 = 18, then 3% of 1,800 = 3 × 18 = 54.

To calculate 2.5% of 1,250: if 1% of 1,250 = 12.5, and 2% = 25.0 and 0.5% = 6.25, then the total = 31.25.

FINDING THE RATE PERCENT

Rate is the comparison between two numbers expressed as a ratio, written as a common fraction. For example, to express what percent of x is y: y:x or y/x, what percent of 20 is 8: rate = 8/20 = 0.40 = 40%.

FINDING THE BASE RATE

The determination of the base rate is often a feature of real-life calculations in business. Business finances will often involve determining a number of rates, from how quickly inventory is being distributed, to comparing spending from month to month or year to year, to salary and benefits increases or decreases. All of these determinations require an understanding of base rate calculations.

Base rates are calculated by creating an equation. For example, to determine what number is 25% of 88: x = 25% of 88, the percentage is changed to a fraction, creating x = 1/4 × 88, then x = 22. If it is desired to determine

A man carries a basketful of organic coffee beans harvested throughout the morning at a coffee plantation in Guatemala City. Guatemala is ranked number one in percentage of its crop rated as highest quality. AP/WIDE WORLD PHOTOS. REPRODUCED BY PERMISSION.

60% of what number is equal to 42: 60% of x = 42, the percentage is changed to a decimal, 0.6x = 42, therefore x = 42/0.6, x = 70.

To apply this method of base rate calculation to a business example: a company keeps track of its sales on a monthly basis. What were the sales for the month of October if the November sales of $14,352 were 115% of October sales? October sales = x; 14, 352 = 115% of x; 14,352 = 1.15x; x = 14,352/1.15 = 12,480.

PERCENTAGE CHANGE: INCREASE OR DECREASE

Whenever a problem is expressed by words such as "is 15% more than," or "is 60% less than," or "has increased by 180%," or "has decreased by 38%," the problem requires a calculation of the percentage change, as either an increase or a decrease.

For example, 44 increased by 25% is what number? The new number will be represented by x, whereby x = 44 + 25% of 44; x = 44 + 1/4 (44); x = 44 + 11; x = 55.

Another example: 90 decreased by 40% is what number? The formula is x = 90 − 40% of 90; x = 90 − 0.4 (90); x = 90 − 36; x = 54.

FINDING THE RATE OF INCREASE OR DECREASE

The rate of change may be expressed as the following equation: rate of change = amount of change / original number. For example, if 40 increases to 46, rate of change = (46 − 40)/40 = 15%.

FINDING THE ORIGINAL AMOUNT

If the quantity after the change in a circumstance is known, the original quantity may be found as follows: 96 is 60% more than what number? In this example, 96 is the number after the increase. Let the original number be x: x + 60% of x = 96; x + 0.6x = 96; 1.6x = 96; x = 96/1.6; therefore, x = 60.

CALCULATING A TIP

A number of studies in recent years have determined that in North America, between 30% and 50% of a typical family's yearly budget for food is spent outside of the home, in restaurants ranging from the typical fast food emporiums of all types to more formal restaurant dining. Tips and tipping are the common terms used to describe the gratuity typically paid to a server in a restaurant for

their assistance to a diner during the course of a meal. The money spent on the tip, which is in addition to the cost of the food and the taxes that may apply to the purchase of a meal, is therefore an important factor in the measurement of the total cost of food spending outside the home.

From the perspective of a server working in a restaurant, the correct calculation of the tip is important because it has a direct impact upon their personal income, as typically the tips earned by a server for their work will constitute an important part of their earnings.

The calculation of a tip involves a percentage-based application, usually related to the total amount of the bill, not including sales tax. It is generally accepted that a 15% tip recognizes good service, while a 20% tip tells the server that the service was outstanding. Tips of less than 10% are treated as an expression of the diner's dissatisfaction with the server and the establishment about the meal.

Assume a 15% tip in the following examples: 15% = 15/100 = 0.15. Where a restaurant bill totals $28.56, without tax being added to the total, to calculate the tip: .15 × 28.55 = 4.28. Thus, the 15% tip on this bill is $4.28.

It is unusual to leave a precise amount such as this for the tip, especially if the bill is being paid by cash. Custom may dictate that if the patron is paying by cash, a rounded figure that approximates the 15% will be left for the server, perhaps $4.25 or $4.50 in this example.

Note that when using the 1% method, this tip could be also calculated as follows: 10% of 28.55 = 2.85; 5% is ½ of 10% = 1.43. Thus, the total is 4.28.

COMPOUND INTEREST

Bank interest is expressed as a percentage. If funds are left in a bank account as a savings, they will attract what is referred to as compound interest, which is interest calculated both on the principal amount as well as the accumulated interest over time.

For example, in Year 1, $10,000 is deposited to a bank account that will pay the depositor 4% per year. The interest earned in Year 1 will equal $400. The interest to be calculated in Year 2 will be calculated on the original $10,000 as well as the Year 1 interest of $400, for a total of $10,400. 4% = 0.04, so Year 2 interest = 10,400 × 0.04 = $416.

Total monies in the account at the end of Year 2 will be $10,816. The 4% rate will apply indefinitely until the money is withdrawn in this example.

RETAIL SALES: PRICE DISCOUNTS AND MARKUPS AND SALES TAX

Many aspects of retail sales advertising are expressed in percentage terms. Sale prices, discounts, markups on merchandise, and all sales tax calculations depend on percentages. The various methods set out below assist in determining the various ways that retail sales are dependant upon percentages.

Discounts and markups: a discount is any sale where the seller claims that the goods are being sold at less than the regular or listed price. In some cases, the original price of the item is known, as is the percentage discount. The sale price is not known and it must be calculated, as follows: A refrigerator was said to have a list or regular price of $625. In the appliance showroom, there is a tag placed on the refrigerator advertising the item as on sale at 40% off its regular price. To find the sale price, 40% = 0.40; 40% of 625 = 0.40 × 625 = 250; 625 − 250 = 375.

In this example, the discount of 40% is $250, and the sale price is therefore $375. As an alternative method for calculating the sale price, 100% − 40% = 60%; 60% = 0.60; 0.60 × 625 = 375.

The next type of discount application commonly required in retail sales is the computing of the percentage discount advertised in any given situation. A used motor vehicle is advertised by its owner as being for sale at a price of $8,500. The advertisement states that the vehicle is worth $12,000 and that it is being sacrificed at the $8,500 price because the owner is relocating to another country to take a new job. The percentage by which the vehicle price is being discounted is calculated as follows:

Percentage discount = original price − sale price / original price × 100%; percentage discount = 12,000 − 8,500 / 12,000 × 100% = 3,500 / 12,000 × 100% = 29.17%.

The opposite concept in retail sales is the notion of the markup. While discounts are typically a part of the retail process that is advertised to the public, the markup is primarily an internal mechanism within a particular retailer.

Items that are sold in retail stores are often manufactured or assembled elsewhere, and they are purchased by the retailer on what is known as a wholesale basis. The ultimate sale price offered by the retail store to a purchaser will be the price paid by the retailer to obtain the item, plus an amount reflecting the relationship between what the retailer paid for an item themselves and what it will be sold for to the public. This amount is the markup. It is also referred to in some businesses as a margin, as in a business operating on a small margin, or the markup may also be described as the gross profit (the profit before costs and overhead is deducted). The relationship

between cost, markup, and the retail or selling price for any item may be expressed in this simple equation: cost + markup = selling price.

Markups will be quoted as either a percentage of the cost price or of the selling price of an item, depending upon what is customary in that particular business. To compute selling price, the following example sets out the process: A hardware store buys drills a cost of $145 per drill. The store marks up the cost 65% based on its cost. The selling price is determined by 65% = 0.65; markup = 0.65 × 145 = $94.25; selling price = 94.25 + 145 = $239.25.

Alternatively, the known markup can be added to 100%, creating a total percentage figure, to perform the calculation: 100% + 65% = 165% = 1.65; selling price = 1.65 × 145 = $239.25.

SALES TAX CALCULATIONS

In most jurisdictions in the world, anyone purchasing consumer goods, ranging from bubble gum to motor vehicles, will be faced with the imposition of a sales tax. Such taxes, depending upon the location, may be imposed by city, state or province, or national governments. Tax rates vary from place to place; it is common to find 5% sales taxes. In some countries what are referred to as goods and services taxes, when combined with existing local taxes, can have a combined impact of 15% or more on a consumer purchase.

When assessing the price of an item offered for sale by a retailer, the total cost of the item must be assessed with the applicable taxes taken into account. For example, a new vehicle dealer is selling a pickup truck for $21,595, plus taxes. If the applicable tax rate is 4.5%, the total cost of the item is 4.5% = .045; tax = .045 × 21,595 = $971.76; total cost = 971.76 + 21,595 = $22,566.78.

Another factor in relation to the calculation of costs is the fact that the retailer may also have paid taxes on their purchase, which are being passed along. For this reason, the actual savings on a discounted item that is purchased has two components: the available discount on the price of the goods in question, and a reduction in the sales tax otherwise applicable to the price.

For example, a television is listed at a regular price of $649. It is then the subject of a "one third discount." The total savings available to the consumer are as follows: Price discount is ⅓ discount = 33.3%; discount = 0.333; discount = 0.333 × 649 = 214.17; discount price = 649 − 214.17 = $432.88.

If the applicable sales tax was 5% the sales tax payable on the discounted price would be tax rate 5% = 0.05; tax on discount price = 0.05 × 432.88 = 21.64; total cost of discounted item = tax + discount price = 432.88 + 21.64 = $454.52.

Had the television been purchased at the regular price, the sales tax would have been taxed at regular price = 0.05 × 649 = 32.45. The total cost of the television at its regular price is 649 + 32.45 = $681.45; total savings on the discounted television purchase is regular price total cost − discounted price total cost: 681.45 − 454.52 = $226.93.

REBATES

A variation on the notion of discounts is that of the rebate. A rebate occurs where a retail business sets a particular advertised or published sale price, and then will offer to refund or discount to the customer a fixed amount or percentage of the sale price. Rebates are frequently advertised in retail sales, and they are most common in the automotive sector, and they are also employed in the sale of various kinds of electronic devices and computer hardware.

For most circumstances, a rebate will have the same effect on a transaction as does a discount: a price that is the subject of a 10% rebate will have the same effect on a transaction as a 10% discount. However, there is one distinction between the impact of a discount and that of a rebate when the rebate is not offered at the retailer, but by way of the format known as a mail-in rebate.

For example, at a computer store that offers various types and brands of computers for sale, a particular computer manufacturer is offering a new computer monitor for sale at a price of $399, less a $50 mail-in rebate. The computer is purchased in accordance with the following transaction: sale price = 399; sales tax rate = 5% = 0.05; sales tax = 0.05 × 399 = $19.95; total cost = 399 + 19.95 = $418.95.

The purchaser is provided with a mail-in rebate card, which sets out the terms of the rebate, namely that upon receipt of the card, the manufacturer will send the sum of $50 payable to the purchaser within 60 days. Therefore, after 60 days, plus the time it takes to deliver the rebate to the manufacturer, the net cost to the purchaser shall be $368.95.

Two percentage-based calculations come into play in this mail-in rebate example. First, the difference is sales tax payable between the mail-in rebate and an identical discount; second, the 60 days or greater that the customer's $50 is out of the customer's control.

SALES TAX CALCULATION: IN-STORE DISCOUNT VERSUS MAIL-IN REBATE

If a $50 discount had been applied to the computer monitor purchase at the time of the transaction, the sale

price would have been reduced to $349, resulting in a total cost to the purchaser of sales tax = 0.05 × 349 = $17.45; total cost = 349 + 17.45 = $366.45.

The difference between the rebate being obtained by the mail-in method and the discount being calculated at the time of purchase at the store is $2.50. To calculate the percentage difference between the total cost of the in store discount purchase and that of the 60-day rebate purchase: rebate cost / discount cost × 100% = percentage difference, or 368.95 / 366.45 × 100% = 1.006%.

To express the cost difference between the in-store discount and the mail-in rebate, the mail-in rebate process is 1.006% more expensive. This calculation as set out here does not place a value on other likely costs, including the time the purchaser would take to complete the rebate form, mail the rebate, and other associated steps required to have the rebate processed.

IMPACT OF THE 60-DAY REBATE PERIOD ON THE COST OF THE PRODUCT

As was noted in the examples dealing with the calculation of percentages, an interest rate measures the value of money over a period of time. Interest rate calculations are useful not only to calculate an increase in the value of money (such as the rate on interest being compounded on money being held in a bank account), but as is illustrated by the 60-day rebate, the interest rate percentage calculation can be used to confirm a loss of value over a period of time.

The calculation of the difference in the total cost of the refrigerator confirmed that the in-store discount total price of $366.45 was $2.50, or 1.006% less than the mail-in rebate total price of $368.95. The next calculation will illustrate what happens to the $50 rebate during the 60-day rebate period.

Assume that if the $50 were placed in a bank account, it would earn interest at a rate of 4% per year. Had the customer purchased the refrigerator by way of an in-store discount, the $50 discount would have been an immediate benefit to the purchaser, deducted at that point from the price paid to the retailer.

By waiting 60 days to receive the rebate (the minimum period, given that as a mail-in rebate there are additional days of mail and processing by the manufacturer), the purchaser lost an opportunity to use that $50 sum. The percentage interest calculation will place a value on that loss of opportunity: loss = value of rebate × number of days rebate not available / length of the year × interest rate; value of rebate = $50; mail-in period = 60 days; year = 365 days; interest rate = 4% = 0.04; loss = $50 × 60 / 365 × 0.04; loss = 50 × 0.164 × 0.04; loss = 0.205.

In this example, the loss of opportunity for the purchaser on the $50 rebate paid to the purchaser after 60 days is a small figure, 20.5 cents. The total difference in cost between the in-store discount purchase and the rebated purchase is the difference in total cost, $2.50, and the loss of opportunity on the $50 rebate, $0.205, for a total of $2.705.

However, as with most retail sales examples using relatively small numbers, it is easy to understand the importance of these percentage calculations where the retail price is 10 or 100 times greater. The percentages do not change, but where the percentages are applied to larger numbers, the potential impact on a purchaser is considerable.

UNDERSTANDING PERCENTAGES IN THE MEDIA

It is virtually impossible to read a news article, whether in paper format or by way of Internet service, that does not make at least one reference to a statistic that is described by way of a percentage. Sports, television ratings, employment, stock prices: all are commonly described in terms a percentage. In the media, it is common for percentage figures to be stated as a conclusion. For example, the income tax rate will be increased by 2.5% next year, for all persons earning more than $75,000 per year.

To properly understand how things such as the consumer price index, the inflation rate, the unemployment rate, and similar issues are reported in the media, it is important to keep in mind the mathematical rules concerning percentages and how they are calculated.

The Consumer Price Indexes (CPI) program produces monthly data on changes in the prices paid by urban consumers for a representative basket of goods and services. Comparisons between prices on a month-by-month basis are useful in determining whether living costs are going up or down. To put it another way, the CPI tells how much money must be spent each month to maintain the same standard of living month to month, as the CPI values the same items to be purchased each period.

The CPI is based upon a sample of actual prices of goods that are grouped together under a number of categories such as food and beverages, clothing, transportation, and housing. Each individual item is priced, and the entire costs of the categories are compared with a selected base period. There are a number of adjustments that are also factored into the calculations, to take into account seasonal buying patterns at holidays and well-known sale periods.

The CPI calculations are made as follows: the base period, representing the time against which the current comparison will be made, is equal to 100, based upon 1990

reference data. Assume that the period to be compared is in November 2005: 1990 base price = $100.00; November 2005 price = $189.50.

The increase in the CPI index from 1990 to November 2005 is 89.5% or (189.50 − 100.00)/100. If the December cost of the consumer basket is 191.10, the increase from the base period of 1990 is 91.10% or (191.10 − 100.00)/100. To calculate the percentage increase between November and December, the following process must be carried out: the November value of 189.50 must be subtracted from the December value of 191.10, for an increase of 1.60% when compared to the 1990 rate. To calculate the percentage change between November and December: 1.6%/ 189.5% × 100% = 0.0084 × 100 = 0.84%. Therefore, there was a 0.84% increase in the consumer price index in this example between November and December.

PUBLIC OPINION POLLS

From time to time, specialist organizations, known as polling companies, will be commissioned to gather data from a segment of the population concerning particular issues. The question asked of the people polled may involve a large national issue, such as whether capital punishment ought to be permitted, or whether the maximum speed limits on national highways should be increased or decreased. In some instances, the polling organization may be hired to obtain the opinions of the public in relation to issues that pertain to a local concern, such as whether a town should permit a casino to be constructed within its boundaries.

The manner in which public opinion polls are carried out is a branch of social science. The methods used by the pollsters in the asking of the questions, the number of people who form the sample upon which calculations are made, and the age and the background of the responders are all factors that may impact upon the answers given to the polling company.

From the perspective of percentages, it is important to appreciate that virtually all such public opinion polls are translated, and reported in the media, as a percentage figure. The meaning to be attached to the percentage quoted as the result of the poll must be examined carefully.

For example, a sample of 4,000 people were asked the following questions: Should cigarette sales in their city be banned completely? Should smoking be banned in every public place in their city? For the first question, the following results were noted: 1,900 said, "yes"; 1,800 said, "no"; 250 were "not sure", and 50 "refused to answer." For the second question, the following results were noted: 2,100 said, "yes"; 1,550 said, "no"; 300 were "not sure"; and 50 "refused to answer."

What are the different ways that the results of each of these questions can be expressed as a percentage? Depending upon how the percentage calculation is used in each case, what answers may be given to each of the questions? The percentage calculation for each answer to question 1 on the ban of cigarette sales is "yes" = 1,900/4,000 = 47.5%; "no" = 1,800/4,000 = 40%; "not sure" = 250/4,000 = 6.25%; "refused" = 50/4,000 = 1.25%.

If the poll was to exclude those who refused to answer the question, and only calculate the responses from people who did answer, the percentages for each answer are "yes" = 1,900/3,950 = 48.1%; "no" = 1,800/3,950 = 345.6%; "not sure" = 250/3,950 = 6.3%.

If the poll were further defined as all respondents who had made up their minds and therefore had a positive opinion on the issue, the formula is "yes" = 1,900/3,700 = 51.35%; "no" = 1,800/3,700 = 48.65%. By taking these steps, the polling company might choose to state this result as "more than 50% of respondents to the poll who had formed an opinion on the question were in favor of a ban on the sale of cigarettes in the city."

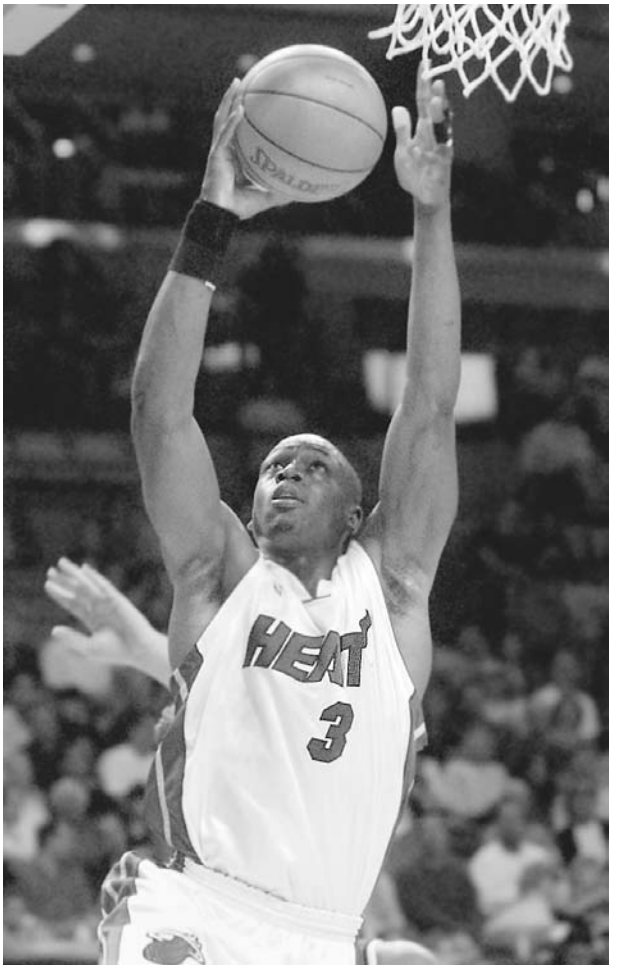
If the poll is defined by who is in favor of the question, the formula is "yes" = 1,900/4,000 = 47.5%; "all other responses" = 2,100/4,000 = 52.5%. The polling company might state this result as "less than 50% of all respondents to the poll stated that they were in favor of a ban on cigarette sales in the city."

The result to the question 2 to ban cigarette smoking in public places generates the following percentage calculations: "yes" = 1,650/4,000 = 41.25%; "no" = 1,550/4,000 = 38.75%; "not sure" = 700/4,000 = 17.5%; "refused" = 100/4,000 = 2.5%.

Using the same analysis as carried out with question 1, if the persons who refused to answer the question are also eliminated from the sample: "yes" = 1,650/3,900 = 42.3%; "no" = 1,550/3,900 = 39.8%; "not sure" = 700/3,900 = 17.9%.

If the persons who were not sure in their answers to the question are removed from the sample: "yes" = 1,650/3,200 = 51.5%; "no" = 1,550/3,200 = 48.5%.

In the same manner as is set out in the question 1 analysis, the manner in which the percentages are calculated in each case can support different conclusions. With the question 2 calculations, when the whole sample of 4,000 answers is examined, only 41.25% of those questioned supported the ban on smoking in public places. By restricting the sample to those with a definitive opinion, a majority of those questioned may be said to support the proposed ban.



Miami Heat's Dwayne Wade goes up and scores against the Atlanta Hawks in the game in Miami. Players are often rated by percentages, such as their field goal percentage. AP/WIDE WORLD PHOTOS. REPRODUCED BY PERMISSION.

USING PERCENTAGES TO MAKE COMPARISONS

It is common in media reports to compare different results in related topics. For example, government spending may be reported in a particular year as having increased 5% over the previous year. The population of a particular state may be stated as having increased by 3% over the past decade.

These calculations are relatively straightforward, because the comparison is being made between single entities, namely a government budget, which would be calculated and measured to be reflected as a total figure, or population, which would have been measured by way of a population count, known as a census.

Percentages are more difficult to put into perspective when they are employed to compare less certain items. For example, if the two public opinion questions and the various answers are compared by way of percentage calculations, the results are not always certain.

In question 1, when only the respondents who had either a yes or a no opinion were calculated, the number of those in favor of the ban on cigarette sales was 51.35%, and those opposed to such a ban was 48.65%. In the question 2 analysis, when only the respondents with a yes or no opinion were counted, the number of those in favor of banning smoking in all public places was 51.5%, those opposed totaled 48.5%.

Based upon the determination of percentage figures that are virtually identical (51.35% and 51.5%) in each question, it would be possible to state the following as a conclusion from the two sets of polling questions, namely a majority of people in the city are in favor of both a ban on cigarette sales and a ban on smoking in all public places.

However, having worked through the calculation to each of the percentages that form the basis of this statement, it is also clear that the use of those percentages in the manner contemplated by this conclusion is not the entire picture. If other parts of the calculation are used to determine a conclusion, it could also be stated that as 47.5% of all respondents were in favor of the ban on cigarette sales, and then a further 41.25% were in favor of the public places ban, the following conclusions are valid: less than 50% of persons polled were in favor of any restriction upon cigarette purchase or usage in the city; a little over 2 out of 5 people polled were in favor of these restrictions.

Percentages and statistics of all types are often stated as a definitive answer or conclusion to an issue. As illustrated in the questions posed above, it is important that the method employed in calculating the percentage be understood if one is to truly understand the significance of the percentage figure that is stated as a conclusion. Where the methodology behind a particular percentage is not stated in a particular media report, the percentage must be regarded with caution.

SPORTS MATH

Another common media report in which percentages are employed in a variety of ways is that of the sports commentary. There are a seemingly limitless number of ways that sport and athletic competition commentaries are enhanced by the use of statistics, many of which are dependent upon percentage calculations.

In the media, there is a recognition that certain statistics go beyond analysis of an individual performance, but are descriptors that convey a definition of enduring excellence. The "300 hitter" is a description applied to a

solid offensive professional baseball player, while a "400 hitter" is in an ethereal world inhabited by legends like Ted Williams and Ty Cobb. A 90% free-throw shooter in basketball has a similar instantaneous public recognition.

The American humorist Samuel Langhorne Clemens, better known as Mark Twain (1835–1910), once said that there are three kinds of lies: % lies, damn lies, and statistics. Whenever a percentage is referenced in a sports article, as with any other media usage of percentages, care must be taken to determine whether the percentage figure being quoted is an accurate indicator of performance, or whether at best it is a lesser or insignificant fact adding only color, and not necessarily insight, concerning the sporting event.

Sports examples of percentage calculation usage are based on daily examples found in the media around the world. For instance, in basketball, an example would be Amanda and Claire as members of their girls' high school basketball team. The coach of the team has been asked to select a most valuable player for the season. While the coach has a personal view of each player based on his assessment of their play through practice and games all season, he decides to do an analysis of their respective offensive statistics. Each player had the following statistics after the completion of the 20-game high school season: Amanda scored 160 total points; 108 2-point shots attempted; 62 2-point shots made; 10 3-point shots attempted; 6 3-point shots made; 21 free throws attempted; 18 free throws made; 17.5 minutes played per 32-minute game. Claire scored 322 total points; 341 2-point shots attempted; 125 2-point shots made; 22 3-point shots attempted; 5 3-point shots made; 81 free throws attempted; 57 free throws made; 28.8 minutes played per 32-minute game. The team scored 887 points on the season.

How can percentages be used to help determine who is having the better season? Conversely, do percentage calculations distort any elements of the performance of these players?

If the 2-point shooting of each player is compared, by calculating the percentage accuracy of each player through the entire season, the following comparison can be made: Amanda = 62 shots made/108 shots attempted = 57.4%. Claire = 125 shots made/341 shots attempted = 36.66%.

The 3-point shooting percentage calculation is as follows: Amanda = 6 shots made/10 shots attempted = 60%. Claire = 5 shots made/22 shots attempted = 22.7%.

The players' free-throw shooting percentages are calculated as follows: Amanda = 18/21 = 85.7%. Claire = 57/81 = 70.4%.

If a newspaper report was written setting out the coach's analysis of the respective play of Amanda and Claire, it is quite possible that such a report might describe Amanda as a better shooter than Claire because her shooting percentages in every area of comparison (2-point shooting, 3-point shooting, and free-throw shooting) are better than Claire's. Conversely, Claire has scored the most points and she has played more minutes per game than Amanda. When those statistics are assessed, the following percentage calculations can be determined: For Amanda, 160 points scored/887 team points scored × 100% = 18% of the team offense. For Claire, 322 points scored/887 team points scored × 100% = 36.3% of the team offense.

Further, Amanda generated her 18% of the team offense while playing 17.5 minutes per game. Claire produced her 36.3% of the team offense while playing 28.8 minutes per game.

There are certain hard conclusions that the coach in this scenario may have reached based upon the percentage calculations that pertain to Amanda and Claire. Amanda is a more accurate shooter in every aspect of the shooting game. It is likely that based upon these percentages, the coach will create opportunities for Amanda to shoot more often next season.

However, as with many applications of the percentage calculation in a sports context, it is important to have more information about the team and the players to give the percentage statistics more context, and to put the percentages into a better perspective. If Amanda is a weak defensive player, her offensive percentages are placed in a different light. If Claire had performed all season known to all rivals as the team's best player, and thus attracted extra attention from opponents, her shooting percentages would be weighed differently.

Baseball statistics may be the most identifiable percentage in sport, usually expressed as a decimal. For example, a strong hitter in the North American professional leagues will be referred to as a "300 hitter," meaning a batsman with an average of 0.300, or a 30%, success rate. This percentage is calculated by the following formula: Number of hits/Number of at bats × 100% = Batting average.

However, as befits a sport that has been played professionally in North America since the 1870s, statistics have grown out of the game, some clear to even the average fan, and some very obscure. A key percentage used to calculate offensive contributions is that of "on base percentage," which measures how often a batter advances to first base by any of the means available in baseball, namely hit, walk, hit by pitched ball, etc. The percentage

is calculated by the following formula: Total number of times on base / Total number of at bats or plate appearances × 100% = On base percentage.

A very intricate set of percentages has made its way into the analytical end of baseball through the work of Bill James. His approach, which he termed sabermetics, is an attempt to use scientific data collection and interpretation methods that employ various types of percentages in different aspects of baseball to conclude why certain teams succeed and others fail.

North American football is also riddled with statistics. One of those measurements is that concerning the most prominent player on the field, the quarterback. How often the quarterback may successfully throw the ball down field is an important statistic, referred to as passing completion. This percentage is calculated by: Number of passes completed/Number of passes thrown × 100%.

However, much like the basketball examples set out above, this percentage on its own is potentially deceiving. A quarterback who throws 80% of his passes for completions, but never throws a pass for a score, is unlikely to be as successful as the 50% passer who throws for 20 touchdowns in a season.

TOURNAMENTS AND CHAMPIONSHIPS

With the rise in the popularity and the sophistication of college sports in the United States, coupled with the impossibility of having hundreds of teams in a given season playing one another head to head, statistical tools were developed to weight the relative abilities of teams that would not necessarily meet in a season, but each of whom would seek selection to an elite end-of-season tournament or championship.

In American college basketball, hockey, and football, tournament selection is made using what is known as the RPI, or ratings percentage index. This interesting and much debated tool is defined in college basketball as follows: RPI = Team winning percentage/25% + Opponents winning percentage/50% + Opponents' opponents winning percentage/25%.

If a team had a record of 16 wins and 12 losses in a season, they would therefore have a team winning percentage of 16 of 57.14%. The team played opponents whose total record was 400 wins and 354 losses. The opponents' winning percentage is 53.05%. These opponents played teams whose winning percentage was 49.1%, the opponents' opponents' winning percentage: RPI = 57.14/25 + 53.05/50 + 49.1/25, which is RPI = 2.28 + 1.06 + 1.96 = 5.304.

A team will typically have a bigger and better RPI if the team combines its own success with an ability to beat strong opponents that have themselves played a strong schedule. Therefore, a team at the end of a particular season that has a lesser record than a rival, but that has played what the RPI determines to be a demonstrably more difficult schedule, may be selected to compete over the team with the better win/loss record. The RPI has a number of nuances that are not the subject of this text, but it is important to understand that the percentage calculation is at the root of any RPI determination.

Percentiles

The percentile is a ranking and performance tool that is closely related to the concept of percentages. A percentile represents a place on a scale or a field of data, providing a rank relative to the other points on the scale. Percentiles are calculated by dividing a data set into 100 groups of values, with at most 1% of the data values in each group.

Percentages can be expressed in any number from 0 to virtual infinity, with either a positive or negative value as circumstance may determine. However, it is commonly accepted that in many applications where a percentage calculation determines a grade or a score in a particular activity, the percentage is expected to be between 0% and 100%. For example, where a school assignment was graded at 17/20, the assignment has a percentage grade of 85%.

In situations where there is a large class of students, it is often desirable to rank them in order of performance. Ranking provides a measure of how a particular student has performed relative to every other comparable student. For example, hundreds of thousands of potential university students in the United States, with many thousands more worldwide, test for the standard Scholastic Aptitude Test (SAT) every year. The SAT is tested at a multitude of test sites, at various times. Each test in a given year is similar, but the exact questions asked on each of the tests will vary. The SAT has a complicated scoring system generating scores from 0 to 1600, and the administrators of the test recognize that assessing students who have taken different versions of the SAT is very difficult. For this reason the percentile ranking becomes important, as it measures where every student stands relative to every other student who took the test.

Determining where an individual students stands relative to everyone else who took the test is a terrific tool with which to assess relative performance. This determination is done by calculating the percentile.

SAT SCORES OR OTHER ACADEMIC TESTING

The percentile grew from the concept of percentages; for that reason, founded upon the concept of 100, and if the data comprising the test results is regarded as a unit of 100, percentile ranking proceeds in bands from 0 to 99, with the 99th band being that that includes the highest score or scores in the sample.

Each percentile in the sample may have more than one score within it. Further, percentiles are not subdivided. For example, there may be as many as 20,000 test scores produced from one round of SAT testing. If eight students scored a perfect 1600 on the SAT, they would each be described as having a result in the 99th percentile even if, say, 10 students with slightly lower scores were also in the 99th percentile. Similarly, if the 55th percentile, representing 1% of all scores from that test, was determined to be all of the scores between 1010 and 1040, all scores within that percentile band would be described as in the 55th percentile.

One formula to calculate the percentile for a given data value is: Percentile = (number of values below x + 0.5)/number of values in the data set × 100%.

As an example, the following is a sample of the shoe sizes for a 12-member high school boys basketball team: Sample: 14, 12, 10, 10, 13, 11, 10, 9, 9, 10, 11, 9. How is the percentile rank of shoe size 12 determined? First, the shoes sizes must be arranged in values smallest to largest, which create this set: 9, 9, 9, 10, 10, 10, 11, 11, 12, 13, 14. The number of values below 12 is eight, and the total number of values in the data set is 12. The formula to express the percentile rank of the value 12 is $(8 + 0.5)/12 \times 100\% = 70.83\%$. The percentile ranking of the value of the size 12 can therefore be expressed as the 1st percentile.

To calculate the percentile ranking of the size 10, there are three identical sizes in the data set. There are three values in the set below 10. The formula would be $(3 + 0.5)/12 \times 100\% = 29.1\%$. The percentile ranking of the value of all three of the size 10s is expressed as the 29th percentile.

It is also common to express a ranking using a broader term. For example, a student may be described as being in the top 20% of their class, or in the top quarter. These expressions are a paraphrasing of the percentiles known as deciles (groups of 10 percentiles) and quartiles (groups of 25 percentiles). Deciles divide the data set into 10 equal parts, and quartiles divide the set into four equal parts.

The 50th percentile, the 5th decile, the 2nd quartile, and the median are all equal to one another.

Final grades in academic courses are typically expressed as a percentage. Even where alternate methods are used to express performance (as with alpha grades A through F), or as a grade point average, each alternative has an equivalency expressed as a percentage. The percentages are then matched to a particular letter grade that has a range of percentages within it. For example, A+ is the equivalent of 90–100%; A is the equivalent of 80–89%; B is the equivalent of 70–79%; C is the equivalent of 60–69%; D is the equivalent of 50–59%; and F is the equivalent of below 50%.

Letter grades function in a similar way as percentiles, in that each grade includes a potential range of percentage scores, and like the percentile, the percentage scores are not ranked within the assigned grade.

Any area of human performance that is subject to ranking will likely employ percentiles as a measuring stick. Topics can be as diverse as the relative rate of obesity in children, ranking increases or decreases in funding rates for hospitals and schools, and comparing the relative safety rates in relation to speed on highways. These are three of the almost limitless ways that percentiles can be used to assist in a ranking of performance.

Potential Applications

The better understanding of a multitude of everyday concepts and activities will be determined, directly or indirectly, by an appreciation of the ability to perform the percentage calculation.

As further examples, percentages play a key role in the following areas:

- Voting patterns and election results: Percentages are used to take the large numbers of persons who may vote in an election, and reduce the figures to a result that is often easier to understand.
- Automobile performance: Octane is a term that is familiar to everyone who has ever used gasoline as a fuel for a vehicle. In general terms, the octane rating refers to how much the fuel can be compressed before spontaneously igniting, an important factor in optimizing the performance of the internal combustion engine. While the public generally associates high octane requirements as required for certain motor vehicle models with more powerful engines and vehicle performance, the octane rating represents the percentage between the hydrocarbon octane (or similar composition) in relation to the hydrocarbon heptane. For example, an 87 octane rating (a common minimum in the United States) represents an 87 percent octane, 13 percent heptane mixture in the fuel.

Key Terms

Fraction: The quotient of two quantities, such as 1/4.

Percentage: From Latin *per centum*, meaning per hundred, a special type of ratio in which the second value is 100; used to represent the amount present with respect to the whole. Expressed as a percentage, the ratio times 100 (e.g., 78/100 = .78 and so .78 × 100 = 78%).

Ratio: The ratio of a to b is a way to convey the idea of relative magnitude of two amounts. Thus if the number a is always twice the number b, we can say that the ratio of a to b is "2 to 1." This ratio is sometimes written 2:1. Today, however, it is more common to write a ratio as a fraction, in this case 2/1.

- Clothing composition and manufacture: Most clothing is sold with a tag or other indication as to its material composition. For example, it is common to see a label on a shirt indicating 65% cotton, 35% polyester, or a sweater marked as 100% wool.
- Vacancy rates: The availability of vacant apartment space in a particular city is of great importance to prospective residents and existing apartment dwellers alike. The vacancy rate is expressed as a percentage to provide interested persons with an indicator as to the relative ease or difficulty to obtain particular types of rental accommodation. Vacancy rates can be viewed as of a particular period (for example, the vacancy rate in Spokane was 1.8% in April), or as a calculation increase or decrease from period to period (for example, the vacancy rate in Toronto fell 0.7% last month).

Where to Learn More

Books

Boyer, Carl B. *A History of Mathematics.* New York: Wiley and Sons, 1991.

Upton, Graham, and Ian Cook. *Oxford Dictionary of Statistics.* London: Oxford University Press, 2000.

Web sites

College Board. "Scholastic Aptitude Test." (March 29, 2005.) <http://www.collegeboard.com>.

NCAA Tournament Selection, 2005. (March 29, 2005.) <http://www.ncaa.com>.

Perimeter

Overview

A perimeter is the boundary of an area or shape. Its measurement is the total length along the border or outer boundary of a closed two-dimensional plane or curve. The origin of the word perimeter comes from the Greek words *peri* (around) and *metron* (to measure).

The application of perimeters in everyday life is widespread when determining a wide range of mathematical problems such as the amount of fencing needed to encompass a homeowner's property; the number of miles of beach property along a lake; and the distance around the equator of Earth.

Fundamental Mathematical Concepts and Terms

One of the simplest equations for solving a perimeter is that of a square or rectangle, which is the sum of its four sides. The general equation for determining the perimeter of a rectangle is $p = 2W + 2L$, where W = width of the rectangle and L is the rectangle's length. Knowing that a rectangle always has four sides with opposite, equal widths and lengths, a rectangle (for example) with length of 4.3 meters (about 14.1 feet) and width of 6.4 meters (21 feet) has a total perimeter length of $p = 2$ (6.4 meters) $+ 2$ (4.3 meters) $=$ 12.8 meters $+$ 8.6 meters $=$ 21.4 meters (about 70.2 feet).

The equation that determines a perimeter of a circle (also known as its circumference) is $p = 2\pi r$ or $p = \pi d$ (where π = approximately equal to 3.14159, r = radius of the circle, and d = circle's diameter and $d = 2r$). As a specific example, a circle with a diameter of 7.5 meters (about 24.6 feet) has an approximate perimeter of $p = \pi$ (7.5 meters) $=$ 3.14159 (7.5 meters) $=$ 23.6 meters (about 77.4 feet). By knowing the shape of a simple figure, such as a triangle, hexagon, square, or pentagon, its perimeter can be easily calculated. More complicated figures, such as an ellipse, need the tools of calculus in order to calculate its perimeter.

A Brief History of Discovery and Development

Archimedes is known to have found the approximate ratio of the circumference to diameter of a circle with circumscribed and inscribed regular hexagons. He computed the perimeters of polygons obtained by repeatedly doubling the number of sides until he reached ninety-six

sides. His method for finding perimeters with the use of circumscribed and inscribed hexagons was similar to that used by the Babylonians (whose civilization endured from the eighteenth to the sixth century B.C. in Mesopotamia, the modern lands of Iraq and eastern Syria).

Real-life Applications

SECURITY SYSTEMS

A physical barrier around the perimeter of a building may stop or at least delay potential intruders from penetrating inside. Such physical barriers include fences, brick or concrete walls, and metal fencing. A well-known outer perimeter barrier surrounds the White House complex in Washington, D.C., which includes very substantial physical fencing, Secret Service agents, and an assortment of television cameras and high-tech sensors. An effective perimeter security system, especially for critically important properties, may include a combination of several physical barriers, an electronic detection system, and numerous manual procedures. A single barrier completely around the perimeter of a protected property could take only a few seconds to penetrate, while multiple barriers will typically take longer to penetrate. Taller and stronger perimeter barriers will further increase the time it takes an intruder to gain entry to a site.

In all cases, in order to effectively secure a property, the physical barrier must completely surround the property's perimeter. As a result, the installers of a perimeter barrier must first measure the number of feet (or meters) in the perimeter. Because of this measurement, these professionals must know the appropriate equations to calculate the perimeter of a square, rectangle, circle, and other shapes. In many instances, numerous equations will need to be combined due to irregular-shaped perimeters around a facility or property. Because of increased risks of terrorism and criminal activities around the world, security that involves total perimeter protection is becoming more popular at governmental, industrial, and commercial facilities such as airports, correctional centers, court houses, entertainment complexes, military bases, and police stations, along with residential homes.

LANDSCAPING

The use of perimeters in landscaping is a common way to design for particular purposes. For instance, commercial properties may use certain plants and shrubs along the perimeter of their facility for the following reasons: to completely isolate the facility from the public; to create a visual separation between the facility and the public; to soften the appearance of streets, parking areas, and other exterior buildings and structures; and to provide summer shade on parking areas.

Defining a landscape's outer boundaries (its perimeter) with respect to the interior buildings, gardens, and other structures and materials often help to create a better visual effect for the entire property. Homeowners with small urban properties, where neighbors live in close proximity to each other, naturally lean toward defining their perimeters with the use of fencing, hedges, shrubs, trees, and other similar structures. These materials are used for such reasons as identification of property lines, privacy, and overall aesthetic beauty. When larger properties are involved, perimeter framing is less used because of fewer concerns for privacy and other such considerations. However, large properties without visible exterior boundaries will often allow such an open area to look more exposed and unfinished—thus detracting from the overall beauty. Simple placement of plantings along the perimeter will make the entire area look more organized and cohesive. Unless privacy, unattractive outside views, or intrusion of wildlife are a concern, most perimeter plantings only need a light planting of trees and shrubs of various densities, sizes, and textures. In all cases, accurate calculations with respect to the total length of the perimeter is essential.

Perimeters are not only used to define the boundary line of a property. Landscaping within a property can also use perimeter-planting when planting around the boundary of a perennial flower gardens, houses, swimming pools, or other such structures. In each instance, the measurement of perimeters is important when designing an outside landscape.

SPORTING EVENTS

Knowledge of the perimeter of various sport fields is important with respect to the watching, playing, and discussing of the games. For example, the perimeter of an American football field (excluding the end zones) is 920 feet (280 m): two lengths of each 300 feet (91 m) and two widths each of 160 feet (49 m). Since each end zone is 30 feet (9 m) long, the perimeter of each end zone is $(30 + 30 + 160 + 160)$ feet $= 380$ feet. Thus, the total perimeter of a football field including the two end zones is 1,680 feet (about 512 m). Playing strategies by coaches and players depend on knowing the exact measurements of a field's perimeter in such sports as football, soccer, tennis, baseball (which can vary depending on the size of the stadium), basketball, and hockey.

BODIES OF WATER

The calculation of perimeters of bodies of water such as lakes and swimming pools is important for many

One side of the perimeter of a farm is marked with a fence. TERRY W. EGGERS/CORBIS.

reasons. Because shorelines are very valuable property with regards to investments, people like to build expensive houses along lakes. Therefore, it is important to accurately measure the perimeter around a lake so, by knowing the length of each house lot, the possible number of total houses built can be figured. This information is very important, for instance, when surveyors and building contractors are first plotting out new lakeside developments.

When first building swimming pools that are to be used for competitions, it is important to know the perimeter of the pool so that the proper number of lanes can be built. For example, the world swimming organization FINA (International Amateur Swimming Federation or, in French, Fédération Internationale de Natation Amateur) states that the official dimensions for pools used for Olympic Games and World Championships are to be of a total length of 50 meters (164 ft) and a total width of 25 meters (82 ft), with two empty widths of 2.5 meters (8 ft) at each side of the pool. With this information, it is easily calculated that an Olympic-sized pool must have a perimeter of 150 meters (about 492 ft) and contain eight lanes, each with a width of 2.5 meters. That is, the total of 25 meters of width consists of 20 meters (66 ft) of lanes (8 lanes $\times$ 2.5 meters per lane $=$ 20 meters) and 5 meters (16 ft) of empty lanes (2 empty lanes $\times$ 2.5 meters per lane $=$ 5 meters).

MILITARY

The United States military has an important need for physical security barrier walls and systems that can protect its ground forces, military fighting assets such as airplanes and tanks, and critical infrastructure assets from hostile actions. These materials are set up around the perimeter of critical structures, soldiers, and materials in order to assure that enemy forces do not penetrate, attack, and destroy such critical personnel and hardware. These perimeter security devices can be simple, portable coaxial cables laid around the perimeter of buildings, properties, or assets, which emit multiple radio-frequency signals. Strategically placed receivers monitor the signals and trigger an alarm when there is a disturbance along the protected perimeter. Other more complex perimeter security devices can be high-technology corrugated metal barriers that can withstand the blast of high-order detonations

and anti-ram barriers that can withstand the repeated assault by enemy tanks and other motorized vehicles.

Potential Applications

PLANETARY EXPLORATION

Perimeter is such a general term within mathematics that its use will always be important for new applications. For example, as mankind ventures further into the solar system, unmanned rovers with portable power supplies, such as rechargeable batteries, may depend on supplementary power generated on stationary landers. As the rover explores a pre-determined area of a celestial body, such as the moons of Saturn and Jupiter, it would return to the central lander to recharge its power supply. This method is very similar to how motorists check their fuel gauge to make sure they are not too far away from a gas station when the arrow points near empty. In such a scenario, aerospace scientists would calculate the straight-line perimeter of maximum exploration for the rover in order to assure that the rover would never venture too far from its power supply. Knowing this maximum number of kilometers, the scientists then keep track of the actual mileage of the rover, most likely within an internal sensor of the rover, to accurately predict when to return to base camp.

ROBOTIC PERIMETER DETECTION SYSTEMS

The U.S. Department of Defense's Defense Advanced Research Projects Agency (DARPA) and Sandia National Laboratories' Intelligent Systems & Robotics Center (ISRC) are developing and testing a perimeter detection system that uses robotic vehicles to investigate alarms from detection sensors placed around the perimeter of protected territories and buildings. Such advanced technologies that involve the use of perimeters allow humans to perform other, more important tasks, and eliminate the loss of human lives from investigating possible intrusions.

Where to Learn More

Books

Bourbaki, Nicolas (translated from French by John Meldrum). *Elements of the History of Mathematics*. Berlin, Germany: Springer-Verlag, 1994.

Boyer, Carl B. *A History of Mathematics*. Princeton, NJ: Princeton University Press, 1985.

Bunt, Lucas N.H., Phillip S. Jones, and Jack D. Bedient. *The Historical Roots of Elementary Mathematics*. Englewood Cliffs, NJ: Prentice-Hall, Inc., 1976.

Web sites

Rores, Chris Rorres. Drexel University. "Archimedes." Infinite Secrets. October 1995. <http://www.mcs.drexel.edu/~crorres/Archimedes/contents.html> (March 15, 2005).

Sandia National Laboratories. "Perimeter Detection." November 4, 2003. <http://www.sandia.gov/isrc/perimeterdetection.html> The Intelligent Systems & Robotics Center. (March 15, 2005).

Thordarson, Olafur, Dingaling Studio, Inc. "Project for an Olympic Swimming Pool, 1998." October 1995. <http://www.thordarson.com/thordarson/architecture/laugardalslaug.htm> (March 15, 2005).

Perspective

Overview

Perspective is the geometric method of illustrating objects or landscapes on a flat medium so that they appear to be three dimensional, while considering distance and the way in which objects seem smaller and less vibrant when they are farther away. The items must be portrayed in precise proportion to each other and at specific angles in order for the effect to be realistic. In art, perspective applies whether the painting or drawing depicts a landscape, people, or objects.

Fundamental Mathematical Concepts and Terms

Basically, perspective works when a series of parallel lines are drawn in such a way that they all seem to head for, and then disappear at, a single point on the horizon called the vanishing point (see Figure 1). The parallel lines running toward the vanishing point are referred to as orthogonals. The vanishing point itself is considered the place that naturally draws the eye in relation to the other objects in the composition, regardless of the size or subject of the work of art, and the horizon is a straight line that splits the image, placed according to the artist's point of view. The higher the artist's vantage point, the lower the horizon appears in the rendering, and vice versa. More than one vanishing point can be applied to a work of art, giving the illusion that the picture bends around corners or has several points of focus. These compositions are referred to as having two-point, three-point, or four-point perspective.

Perspective is based upon the assumption that one is viewing the image from a single point, and is therefore, sometimes referred to as centric or natural perspective. It is also possible to examine three-dimensional space from two points, the study of which is known as bicentric perspective.

A Brief History of Discovery and Development

Early paintings and drawings, prior to the invention of perspective, tended to appear flat and out of proportion. They lacked a sense of realism. Linear perspective, the first method of creating art that was more precise in its portrayal of its subjects, was invented by Filippo Di Ser Brunellesci (1377–1446), a sculptor, architect, and engineer in Florence, Italy. Brunellesci was responsible for building several of Florence's most famous structures including the Duomo (dome of the main cathedral) and

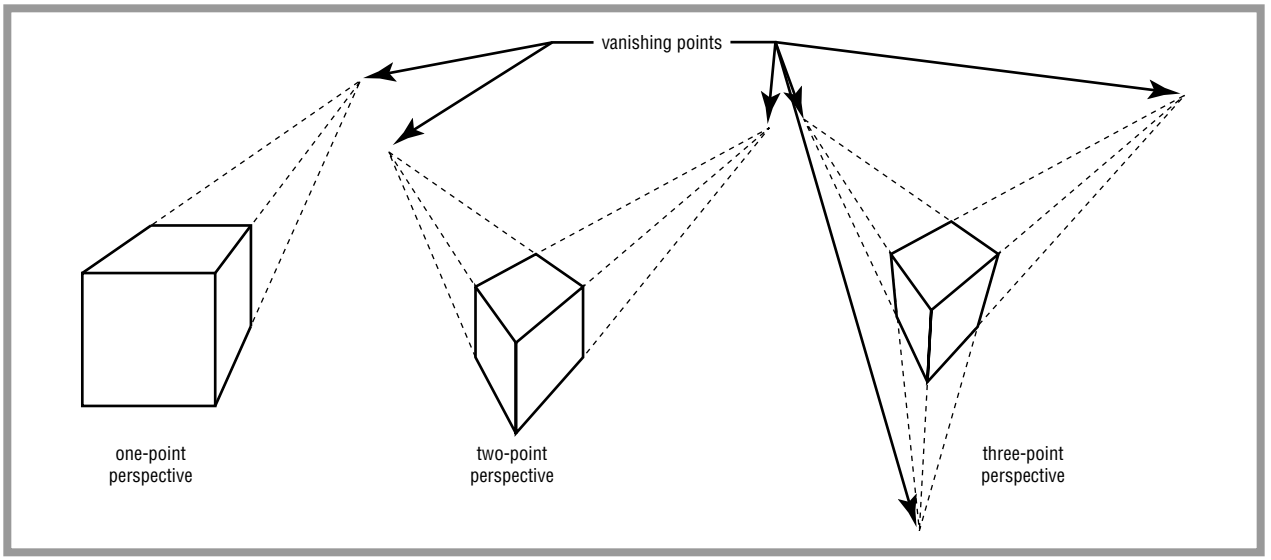


Figure 1.

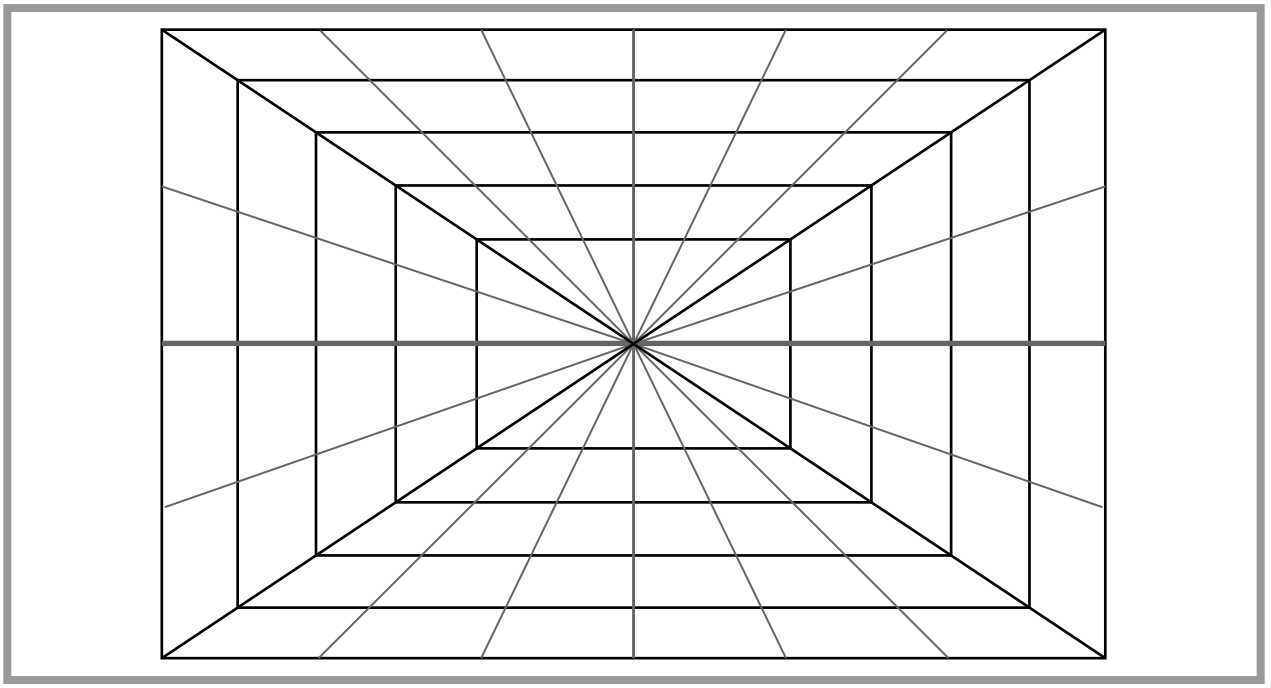


Figure 2.

church of San Lorenzo. Brunellesci experimented with creating a single line of sight, toward a vanishing point, by viewing a reflection of a picture or image through a peep hole in a sheet of paper and thereby focusing his vision on a single line (see Figure 2).

Brunellesci never recorded his findings, but may have passed them on to other artists and architects through demonstrations or word of mouth. The first written account of the use of perspective was recorded by the Italian architect Leon Battista Alberti (1404–1474),

who initiated the use of a glass grid through which the artist would look at the subject while painting in order to assist in creating the proper perspective. Alberti determined that he could use a geometric technique in order to mimic what the eye saw, and also that the distance from the artist to the scene being painted had an effect on the rate at which the image appeared to recede. Alberti said that the artist created a sort of visual pyramid, turned on its side, between himself and the painting, where his line of sight connected to the vanishing point on the work of art. The surface of the painting itself was the base of the pyramid and the painter's eye formed the summit. Alberti considered it necessary to maintain that position in order for the artist to accurately capture the perspective of his subject on the canvas.

The first surviving example of the use of perspective in art is credited to Donato di Niccolò di Betto Bardi (1386–1466), more commonly known as Donatello, an Italian sculptor during the early part of the Renaissance. Of his surviving work, most prominent are sculptures he created for the exterior of the Florentine cathedral, including St. Mark and St. George. The latter is a marble relief that depicts Saint George killing the dragon, and the work shows some indication that Donatello attempted to use perspective within the scene. Some of the lines used to create the illusion were most likely inaccurate, as the perspective is less than perfect, so it cannot be said for certain that he was applying this then-new methodology.

However, in later works, it becomes more obvious that Donatello was aware of the principles of perspective. In a bronze relief panel he designed for the font at the Siena cathedral, titled *Feast of Herod*, Donatello clearly utilized a vanishing point and orthogonals. While there is a slight imperfection in the panel, in that the orthogonals do not meet precisely at the same point, it is likely this defect was not part of the original sketches, but instead resulted at some point during the execution in bronze.

Masaccio (c. 1401–1428), considered with Donatello and Brunellesci to be among the founding artists of the Italian Renaissance, showed no signs of attempting to use perspective in his first known painting, *Madonna and Child with Saints*. However, his three most famous works painted near the end of his life all use linear perspective. One of these, *Trinity*, which was done for Saint Maria Novella in Florence, is thought the oldest perspective painting to still survive today. It depicts the crucifixion of Jesus Christ, with key figures such as John the Baptist and the Madonna framing him in a pyramid fashion, and God hovering above. Masaccio supposedly discussed "Trinity" with Brunellesci. The work itself was painted based on a strict grid that was applied to the surface before any painting began. Every detail is in precise perspective, down to the nails holding Christ to the cross. In another perspective painting, "Tribute Money," Masaccio used linear perspective not just to create a realistic portrayal of the scene from the lives of St. Peter and St. Paul, but also to direct the viewer's eye in such a way that the painting becomes a narrative. Christ stands in a group of his followers, and it is his head that is the vanishing point on which the viewer focuses.

The advent of the camera obscura in the mid-fifteenth century offered another way to examine perspective. Based on similar techniques as the peephole experiments, the camera obscura allowed light into a darkened room through a small hole. An image was then projected onto a wall and the artist attached paper to the surface in order to trace it. The act of tracing guaranteed the artist would achieve the proper angles and proportions of perspective.

Other artists went on to do additional experiments in perspective, and to perfect the technique. Leonardo da Vinci, noted as an artist, inventor, and mathematician, did much to further the understanding of how perspective applied to distance, shape, shadows, and proportion in art. He was the first artist to work with atmospheric perspective, where the illusion of distance was created through using fainter or duller colors for objects meant to be farther from the viewer. By combining this knowledge with other mathematical references, such as the standard proportions of the parts of the human figure, he was able to create compositions that appeared realistic and natural. Albrecht Dürer, a noted German Renaissance artist and print maker, experimented with using tools to assist in attaining proper perspective, and kept detailed records of his discoveries. In 1525, he wrote a book in order to teach artists how to represent the most difficult shapes using perspective.

During the seventeenth century, Dutch artists were particularly known for their exemplary use of perspective in their paintings. Pieter de Hooch and Johannes Vermeer were two such painters renown for including such details as floor tiles, elaborate doorways, and multiple walls incorporating perspective in order to achieve the most realistic effect.

Real-life Applications

ART

Artists display the most obvious need for a clear understanding of perspective in their work. In order to

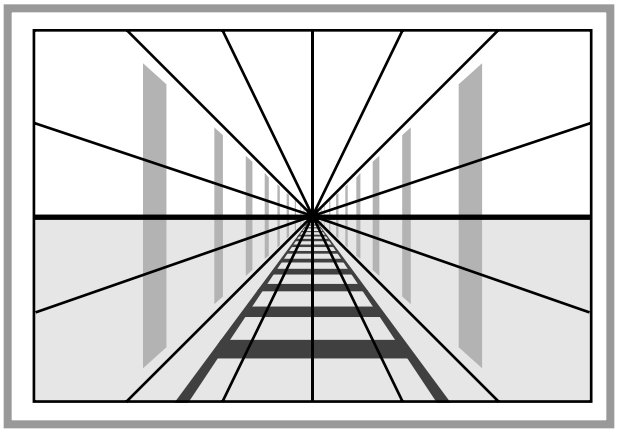


Figure 3.

fashion any realistic depiction of a scene, whether in a simple sketch or a detailed painting, an artist must use the rules of perspective to guarantee that the proportions and angles of the images appear three-dimensional. Landscapes particularly require exact application of perspective in order to give the illusion of depth and distance. A common illustration of this technique (see Figure 3) depicts a train track heading toward the horizon, the parallel lines of its rails appearing to become closer together as they grow farther away, until they eventually converge at the vanishing point. The picture becomes more complex if the artist wishes to add something along the side of the train tracks, such as trees or telephone poles. Although the artist knows the phone poles must appear smaller as they grow more distant, he needs to determine at what rate their size decreases. By applying the rules of perspective, the artist may sketch in the orthogonals, the diagonal lines that stretch from the vanishing point to the edge of the paper, in order to provide a guideline for the heights of the poles as they gradually shrink into the distance.

This method can be applied to any number of subjects that may appear in a painting, such as a row of buildings that reaches to the skyline or clusters of people scattered across a large room for a party. Orthogonal lines can be placed at any height in relation to other subjects so that smaller objects remain in proportion to larger ones, regardless of their placement in the scene. If a man who is six feet tall stands next to a child who is only three feet tall, the child will appear half the height of the man if they are sketched at the front of the painting or back near the horizon, even though the actual size of each will be adjusted to represent their placement in the composition.

Perspective helps artists render drawings that include buildings much more accurately, as well. If an artist wishes to paint a landscape that includes a house and a barn that are situated at an angle, with the corners of the buildings facing the viewer, perspective allows him to draw the edges of the buildings and their roofs at the correct angles. The horizontal lines that form the top and bottom edges of the buildings, as well as the horizontal lines for the door and windows—if extended straight out to the side—should eventually intersect at a vanishing point. The slanted lines that form the side edges of a pitched roof will also intersect in the same way. If the painting includes a split-rail fence around the farmland, the rails must all angle so that the lines would extend to a vanishing point. In these types of landscapes, the artist will frequently use two-point or three-point perspective in order to set the angles for the different sides of the buildings.

Artists often use the vanishing point as a focal point when composing the layout of a painting. If several people are depicted, it is common for an artist to have their attention directed toward the vanishing point. A person gesturing with an arm might likewise be indicating something at the vanishing point.

ILLUSTRATION

One specific application of artistic talent, illustration, provides books and other publications with artwork to accompany the text. Children's books are a prime example of this, and the simplicity of many of the pictures that illustrate children's stories does not preclude the need to apply perspective to the composition. A child will notice if a picture seems out of proportion, just as an adult will, and as the illustrations carry much of the weight of the storytelling for pre-readers, it is important that everything is rendered correctly and in proportion.

Comic books or graphic novels are other examples of illustration as an art form. As with picture books for children, comic books rely heavily on the pictures to tell the story, with only a small amount of narrative and dialogue to move the plot forward. Each panel of a comic is drawn in perspective, with the occasional pane drawn in such a way as to indicate the action happens in the foreground and is therefore, more important. Using perspective for emphasis allows comics to convey heightened emotion and action in a relatively small space.

ANIMATION

Animation, an art form unto itself, would not be possible without perspective, as the figures would appear flat and lifeless on the screen despite their ability to move. Early animated films were hand drawn a single frame at a

Art and Mathematics—Perspective

Perspective provides flat, two-dimensional works of art with the means to appear three dimensional and realistic. No painting, sculpture, or frieze can seem to have depth or illustrate distance from the viewer if the artist fails to apply the rules of perspective to the composition. In reality, the curvature of the planet combines with the eye's ability to look into the distance and creates the visual effect of perspective where lines appear to converge upon a single point, even when the lines never actually meet, as is the case with the two rails of a train track. This trick of the eye, or perspective, must be replicated as an optical illusion on a flat canvas in a painting in order for it to considered a precise representation of the three-dimensional view seen in real life.

A student of art must learn to apply perspective to whatever he is attempting to create. This holds true of paintings done from life and those created solely from the imagination. While it is possible to sit at an easel and recreate the landscape just beyond the top of the canvas, it is more difficult to create an accurate rendering when the subject is not visible. For this reason, art students learn the principles behind the illusion of perspective. An artist can sketch a horizontal line onto a canvas and create both horizon and vanishing point, then add orthogonal lines to assist in creating an accurate, realistic landscape, even in a room without a view.

time, and the precise measurements required to achieve perfect perspective made it easier for the artist to recreate the background of the film over and over, while limiting variances that might have made the finished film appear inconsistent or fake.

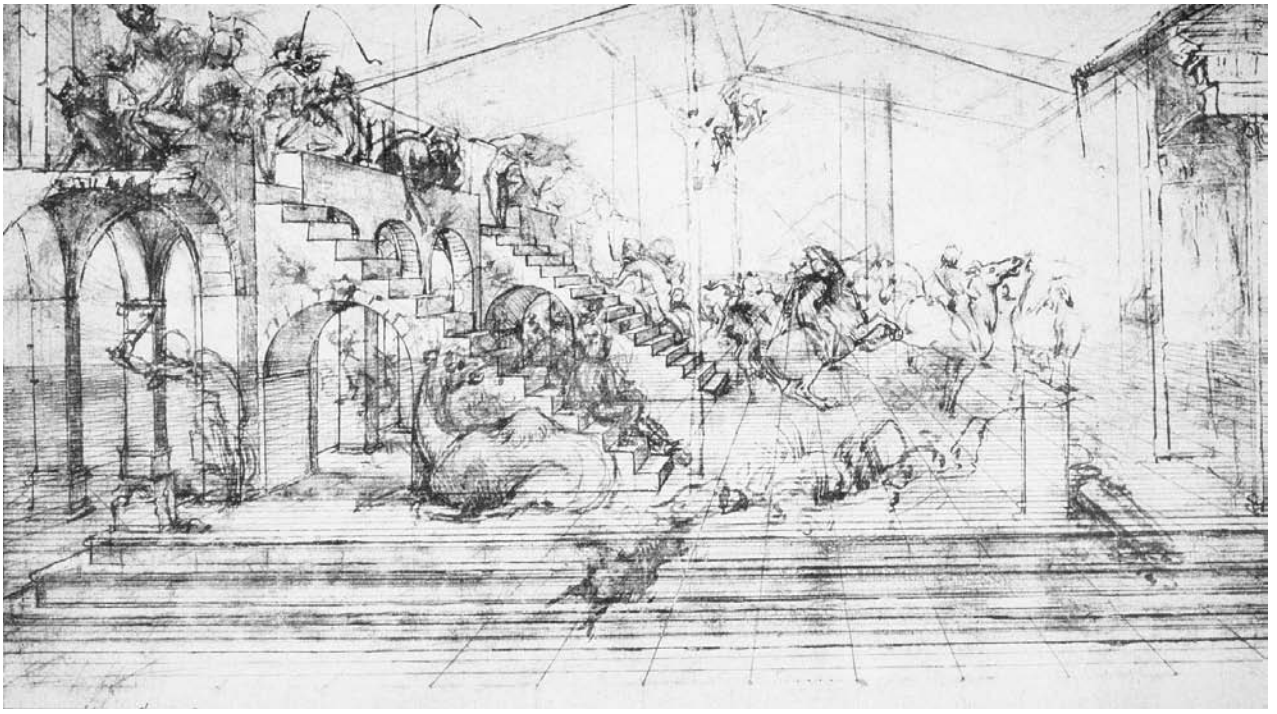
As animation has grown more technical and the art has shifted from paper to computers, it has become more important that the angles and lines required to give the illusion of a three-dimensional setting remain constant. Animators can now feed mathematical calculations into a computer where a graphics program will plot the coordinates for the horizon and the vanishing point. Once this information is computerized, it is saved in the machine's memory and applied whenever that particular background is needed for the film. The computer software allows the animators to program shifts in shadow based on time of day or night for the story, to alter the camera angles, or even to add in new background structures such as a new building or taller trees due to the passage of time. The changes are made automatically within the parameters of the perspective already programmed into the computer.

One modern example of the use of this technology is the Walt Disney Company's film *Beauty and the Beast*. This animated movie applied new technology to centuries-old theories of perspective to create a scene where the Beast and Belle dance in an animated virtual reality ballroom. The scene consists of a large ballroom with rounded walls and a tiled floor, and the film gives the illusion of a living couple twirling around the dance floor as the camera pans around them. The animators programmed the computer to maintain the proportions of the room, with the apparently rounded backdrop, and the tiles on the floor decreasing in size as they grew more distant from the camera. As the animated couple dances and the camera follows them, the vanishing point is required to shift with each movement so that it will remain steady in relation to the eye of the audience and the illusion of depth may be maintained.

FILM

Animated films are not the only ones concerned with perspective. As live action films include more and more special effects that require actors to perform in front of green screens or blue screens, perspective becomes the concern of special effects artists. Obviously the effects artists need to apply perspective when generating the background, as they would with an animated film, but in addition they must maintain the size ratio between the live actors who will be part of the finished scene and any computer graphics components, including scenery and creature effects. The actors must also perform in relation to special effects that are not present while they are filming. While stand-ins are sometimes utilized, it is also helpful to apply the same lines of perspective that an artist would use when composing a painting. An actor might address himself toward what will end up being the vanishing point of the scene, allowing the special effects artists to fill in the graphics around the same point, creating the illusion that all of the components of the film actually took place at the same time.

An example of combining live action with digital backgrounds is the film version of the Frank Miller graphic novels, *Sin City*. In this film, the actors performed their scenes against a green screen, often without even the benefit of another actor to whom they could address their lines. The background, a heightened noir-style city in stark black and white, was created on the computer using a three-dimensional digital program. Using the graphic



Study for perspective with animals and figures by Leonardo da Vinci. BETTMANN/CORBIS.

novel as a template, the director recreated the look and feel of each panel of the comic by mimicking the perspective of each shot. The background maintained the perspective and all of the angles from the original source material, and the actors were placed in relation to that background to make it seem as if the graphic novel itself had come to life.

Another optical illusion popular in film—particularly fantasy or science fiction films—is making actors of similar heights appear vastly different in size. *The Lord of the Rings* trilogy faced this challenge when the filmmakers attempted to create a world shared by several species of varying heights. When an actor playing a short Hobbit filmed a scene with an actor playing a normal sized person, it was not only necessary to have the actors appear to be different heights. The sets around them also had to be altered so that items that appeared average size for the man would be oversized for the Hobbit. Props, such as a ring or a mug of ale, could be duplicated in varying sizes and then substituted for each actor according to their character's size, but the background and furnishings were more complicated. The set designers used perspective to determine the precise proportions for each item and then used forced perspective filming in order to create the optical illusion that the two actors were actually using the same items. For example, in a scene where the wizard, Gandalf, and the Hobbit, Bilbo, are seated at a long table, the front of the table was cut down to be smaller than normal, so that Gandalf would appear to be cramped. The back half of the table was sized normally so it would appear to fit Bilbo. Items placed on the table at the joining point helped disguise that the table was not all one size, and the camera was placed at an angle to shoot down the table's length, taking advantage of the fact that perspective would help make it seem to grow smaller at a distance. The actors themselves stood several feet apart, but staring straight ahead, and were filmed in profile to give the illusion of their facing each other. Perspective made the more distant actor playing Bilbo appear smaller than the actor closer to the camera.

INTERIOR DESIGN

Interior designers and decorators are responsible for the layout and design inside a house, and frequently use perspective as a tool to maximize the potential of a living space. An architectural detail such as exposed beams—which were originally solely a functional aspect of a house, used to brace walls and support the roof—can make a room appear to be longer than it really is. Looking carefully at the beams running parallel to each other, they seem to grow closer together as they move toward

Leonardo da Vinci's "Window" for Recording Proper Linear Perspective in Art

Italian artist, inventor, and mathematician Leonardo da Vinci (1452–1519) understood that linear perspective was necessary in order for a painting to appear realistic. In order to practice transposing the exact lines and angles of the world as he saw them, Leonardo began to use a window as a framework. When he looked out the window, whatever he saw became the subject of his painting, as if the edges of the window were the edges of a canvas. He would then attach a piece of paper to the window so that the natural light shone in from outside and he was able to see the outline of the scene through the paper. It was necessary for him to cover one eye when working, so that he would, in effect, be looking at the three-dimensional world from a two-dimensional viewpoint. He would then go on to trace what he saw through the window onto the paper. Leonardo da Vinci accurately captured all of the lines of perspective as they appeared in nature. This exercise enabled him to learn how perspective affected the composition. He discovered that his own distance to the window, as well as the distance of the objects outside to the window, changed the perspective of the scene. If he shifted to the left or the right, the vanishing point on the horizon also shifted on his paper. It was also possible for Leonardo to sketch in guiding lines, orthogonals, to help him maintain the size ratio between various items in the composition, regardless of where they appeared in relation to the vanishing point. Leonardo proceeded to apply what he learned to his painting. Early sketches of his work illustrate how he composed his work to include a vanishing point that was logical in relation to the subject of the painting.

The famous painting, *The Last Supper*, clearly illustrates Leonardo da Vinci's use of perspective. While the scene itself shows only minimal depth, concentrating more on the length of the dining table as it stretches the width of the painting, with Christ and his disciples positioned along the back, Leonardo applied his knowledge of perspective to create the rear walls of the room. Jesus himself, seated at the center point between his followers, provides the focus of the painting, and his head serves as the vanishing point on the horizon for the composition.

the opposite end of the room from the viewer, just as train tracks seem to converge toward a vanishing point when viewed from a distance. In a house, the beams reach the supporting wall before they appear to meet each other, but the vanishing point still exists. If one could see through the wall and extend the beams indefinitely, they would illustrate a textbook example of perspective. As it stands, the optical illusion they create gives a home a more spacious feel. Anything that adds horizontal lines to the overall look of a room—tiles or hardwood flooring, a chair railing or molding, decorative detail on a ceiling, built in bookshelves that run the length of a wall—gives the impression that a room is longer and more spacious.

A similar illusion that also uses perspective to make a room seem larger is adding a large mirror to a wall. If an entire wall contains a mirrored surface, it will seem to double the size of the room by reflecting it back upon itself. By staring into the mirror, a viewer will notice that the reflected walls seem to angle inward, just like the train tracks in a perspective painting. The illusion of additional space suddenly looks more like the view out a window than an addition to the room. The mirror effect is particularly popular when a designer can place it opposite a window, thereby reflecting not only additional space from the room, but the light and the view from outside as well, creating an open effect.

Another decorating effect that makes use of perspective is the artistic treatment known as *trompe d'oeil*. Literally meaning "trick of the eye," this painting technique involves rendering a highly realistic looking painting or mural directly onto the wall of a room in an attempt to make it appear completely authentic to the viewer. In some cases, the painting is something simple, such as a statue on a pedestal standing in an alcove. Someone looking at the painting from a distance will be tricked into believing that the wall really does curve back at that point, and that the piece of art in question is actually a three-dimensional statue. Only when they draw nearer will they realize that the statue is painted on the wall. The artist uses lines of perspective to create the illusion, perhaps giving the alcove portion of the painting a tiled pattern or gradually lightening the tone of the paint used since colors fade at a distance, all in order to make the wall seem to curve.

Key Terms

Bicentric perspective: Perspective illustrated from two separate viewing points.

Centric perspective: Perspective illustrated from a single viewing point.

Orthogonals: In art, the diagonal lines that run from the edges of the composition to the vanishing point.

Vanishing point: In art, the place on the horizon toward which all other lines converge; a focus point.

Other examples of the use of *trompe d'oeil* may include a painted window or doorway, including the view through that opening. Perspective is applied as it would be in any landscape, so that the view through the painted window or door mimics what one might see through an actual hole at that point, or else the artist might create an entirely imaginary landscape, giving a city apartment the luxury of a view of the beach or the countryside.

Trompe d'oeil may also be applied to an entire wall, as in a mural. This sort of effect can involve multiple illusions, depending on the images chosen for the composition. Some of the wall might be painted as if it were still part of the house, with the rest providing some sort of outdoor view. Examples might include a painting of a balcony that overlooks the garden, with the majority of the perspective applied to the images that are meant to be more distant, and other, more subtle techniques used for the supports of the balcony that are meant to be much closer. However, the lines of the balcony must remain in harmony with the lines of the view, maintaining the same vanishing point, in order to maintain the overall effect.

LANDSCAPING

Landscapers and landscape architects do for the outdoors what interior designers do for the inside of a building. By applying the rules of perspective when laying out a garden, park, or other property, landscapers can make a small piece of land seem larger or grander than it might otherwise appear. A building with a straight driveway can be made to appear farther from the road by planting a series of trees along each side of the drive. The effect is similar to that of a painting of a road with trees lining it, the road converging on the vanishing point and the trees shrinking into the distance. Likewise, details such as long, narrow reflecting pools, hedges, stone walls, flower beds, and flagstone or brick pathways help draw the eye in a particular direction and direct the visual focus of the landscape in whatever way the designer sees fit.

Potential Applications

COMPUTER GRAPHICS

Any work done with computer graphics can make use of the rules of perspective. Programs that allow images to appear on the computer screen in three dimensions apply to a range of work, including architecture, city planning, or entertainment.

Architects and engineers can use preprogrammed angles of perspective to create virtual images of buildings or bridges or other large-scale projects, enabling them to test the effect of the new construction in its intended setting without having to build detailed models. City planners can in turn use perspective to get an accurate idea of the layout of a town from the comfort of a desk. Streets and traffic flow, how roads converge, where traffic lights might be most effective, entrances to major thoroughfares, and placement of shopping or public facilities, all may be programmed into a computer and illustrated in a realistic, three-dimensional layout.

Computer game designers can apply perspective to their creations, enabling enthusiasts to enjoy the most realistic experiences possible when playing their games. Accurate perspective can enhance a variety of games, such as those where the participant drives a racecar, pilots an airplane, or maneuvers a space ship through an asteroid field in a faraway galaxy. Likewise, games that involve role play or character simulation can provide realistic settings, such as towns or the interiors of buildings.

Where to Learn More

Books

Atalay, Bulent. *Math and the Mona Lisa: The Art and Science of Leonardo da Vinci*. Washington, D.C. Smithsonian Books, 2004.

Parker, Stanley Brampton. *Linear Perspective Without Vanishing Points*. Cambridge, MA: Harvard University Press, 1961.

Woods, Michael. *Perspective in Art*. Cincinnati, OH: North Light Books, 1984.

Periodicals

Ashcroft, Brian. "The Man Who Shot Sin City." *Wired*. April, 2005.

Web sites

Dartmouth College Web site. "Geometry in Art and Architecture." <http://www.dartmouth.edu/~matc/math5.geometry/unit11/unit11.html/> (April 8, 2005).

Disney's Beauty and the Beast: Unofficial Pages. "Gallery of Key Scenes." <http://www.ilhawaii.net/~beast/pages/bb_gllry.htm> (April 8, 2005).

Drawing in One Point Perspective. Harold Olejarz. <http://www.olejarz.com/arted/perspective/> (April 8, 2005).

Leonardo's Perspective. <http://www.mos.org/sln/Leonardo/LeonardosPerspective.html> (April 8, 2005).

Perspective from MathWorld. <http://mathworld.wolfram.com/Perspective.html> (April 8, 2005).

Wired News. "Sin City Expands Digital Frontier." Jason Silverman. April 1, 2005. <http://www.wired.com/news/digiwood/0,1412,67084,00.html> (April 8, 2005).

Other

The Lord of the Rings: The Fellowship of the Ring. Extended Edition DVD special features. New Line Home Entertainment, 2002.

Photography Math

Overview

Photography, literally writing with light, is full of mathematics even though modern auto-exposure and auto-focus cameras may seem to think for themselves. Lens design requires an intimate knowledge of optics and applied mathematics, as does the calculation of correct exposure. When mastered, the mathematics of basic photography allow artists, journals, and scientists to create more compelling and insightful images whether they are using film or digital cameras.

Fundamental Mathematical Concepts and Terms

THE CAMERA

In its simplest form, the camera is a light-tight box containing light sensitive material, either in the form of photographic film or a digital sensor. A lens is used to focus light rays entering the camera and produce a sharp image. The amount of light striking the film and sensor is controlled by shutter, or curtain that quickly opens and exposes the film or sensor to light, and the size of the lens opening, or aperture, through which light can pass.

FILM SPEED

The speed of photographic film is a measure of its sensitivity to light, with high speed films being more sensitive to light than low speed films. Film speed is most commonly specified using an arithmetic ISO number that is based on a carefully specified test procedure put forth by the International Organization for Standardization (ISO), for example ISO 200 or ISO 400. Each doubling or halving of the speed represents a doubling or halving of the sensitivity to light. Thus, ISO 400 speed film can be used in light that is half as bright as ISO 200 speed film without otherwise changing camera settings. Some films, particularly those intended professional photographers or scientific applications, also specify speed using a logarithmic scale that is denoted with a degree symbol (°). Each logarithmic increment represents an increase or decrease of three units corresponds to a doubling or halving of film speed. ISO 400 film as a logarithmic speed of 27° but ISO 200 film, which is half as fast, has a logarithmic speed of 24°.

Photographic films are coated with grains of light-sensitive silver compounds that form a latent image when exposed to light. Film speed is increased by increasing the size of the silver grains, and the grains in high speed films can be so large that they produce a visible texture, or

graininess, in photographs that many people find distracting. Therefore, photographers generally try to use the slowest possible film for a given situation. In some cases, however, photographers will deliberately choose a high-speed film or use developing methods that increase grain in order to produce an artistic effect. The choice of film speed is also affected by factors such as the desired shutter speed and aperture.

Camera lens. UNDERWOOD & UNDERWOOD/CORBIS.

LENS FOCAL LENGTH

The focal length of a simple lens is the distance from the lens to the film when the lens is focused on an object a long distance away (sometimes referred to as infinity, although the distance is always finite), and is related to the size of the image recorded on the film. Given two lenses, the lens with the longer focal length will produce a larger image than the lens with the shorter focal length. Most camera lens focal lengths are given in millimeters. A lens with a focal length of 100 mm (3.9 in) is in theory 100 mm (3.9 in) long, but camera lenses consist of many individual lens elements designed to act together. Therefore, the physical length of a camera lens will not be the same as the focal length of a simple lens. Zoom lenses have variable focal lengths, for example 80–200 mm (3.1–7.9 in), and also variable physical lengths. The physical lens length will also change as the distance to the object being photographed changes.

Lenses are often described as telephoto, normal, and wide angle. Normal lenses cover a range of vision similar to that of the human eye. Wide angle lenses have shorter focal lengths and cover a broader range of vision whereas telephoto lenses have longer focal lengths and cover a narrower range of vision. All of these terms are relative to the physical size of the film being used. A normal lens has a focal length that is about the same as the diagonal size of the film frame. For example, 35 mm (1.4 in) film is 35 mm (1.4 in) wide and each image in a standard 35 mm (1.4 in) camera is 24 mm (0.9 in) by 36 mm (1.4 in) in size. The Pythagorean theorem can be used to calculate that the diagonal size of a standard 35 mm (1.4 in) frame is 43 mm (1.7 in). Lenses are usually designed using focal length increments that are multiples of 5 mm (0.2 in) or 10 mm (0.4 in) and 40 mm (1.6 in) lenses are not common so, in practice, the so-called normal lens for a 35 mm (1.4 in) camera is a 35 mm (1.4 mm) or 50 mm (2.0 in) lens. Manufacturers of cameras with film sizes or digital sensors of different sizes will sometimes describe their lenses using a 35 mm (1.4 in) equivalent focal length. This means that the photographic effect (wide angle, normal, telephoto) will be the same as that focal length of lens used on a 35 mm (1.4 in) camera.

SHUTTER SPEED

The amount of light striking the film is controlled by two things: the length of time that the shutter is open (shutter speed) and the lens aperture. Shutter speed is typically expressed as some fraction of a second, for example 1/2 s or 1/500 s, and not as a decimal. Manual cameras allow photographers to choose from a fixed set of mechanically controlled shutter speeds that differ from each other by factors of approximately 2, and the shutter is opened and closed by a series of springs and levers. For example, 1/2, 1/4, 1/8, 1/15, 1/30, 1/60, 1/125, 1/250, and so forth. Note that the factor changes slightly between 1/8 and 1/15, and then again between 1/60 and 1/125. In order to make the best use of limited space on small cameras, film speed dials or indicators in many cases use only their denominator the shutter speed. Thus, a camera dial showing a shutter speed of 250 means that the film will be exposed to light for 1/250 s. Electronic cameras, whether film or digital, contain microprocessors and can offer a continuous range of shutter speeds. The shutter speeds can be set by the photographer or automatically selected by the camera.

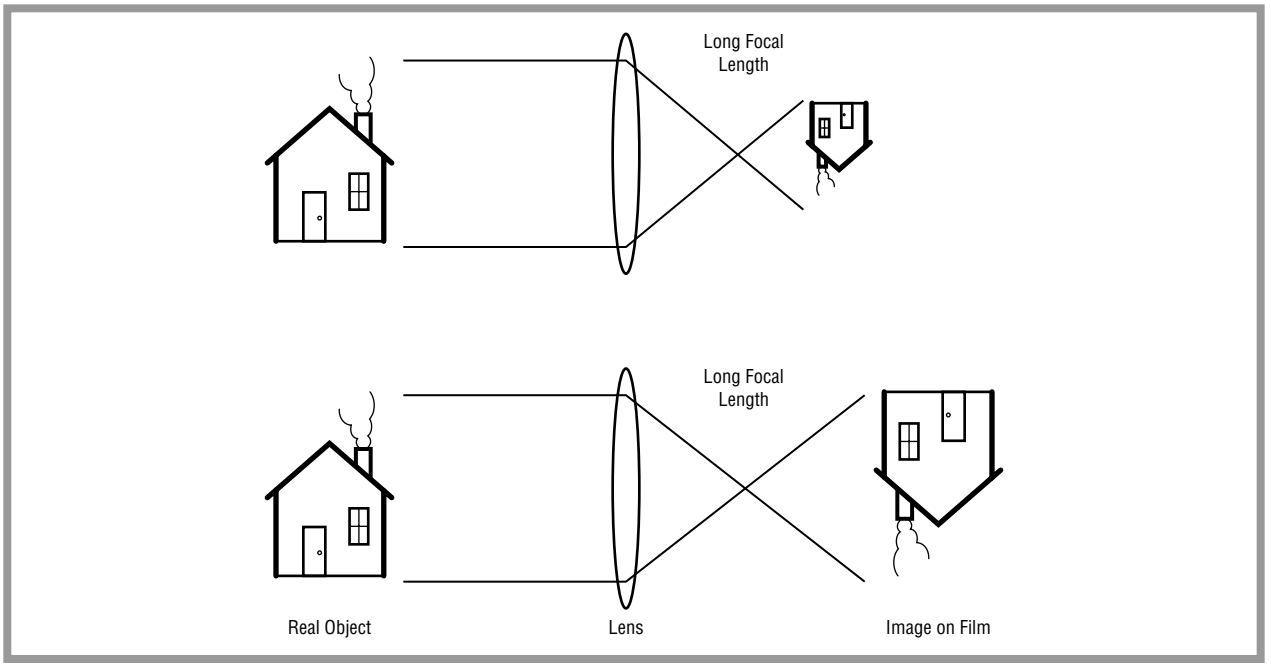


Figure 1.

LENS APERTURE

Lens aperture is the diameter of the opening through which light passes on its way to the film. The larger the aperture, the wider the opening and the more light that will pass through the lens to the film. Aperture is expressed as a so-called f-stop or f-number that is the quotient of the lens focal length divided by the diameter of the aperture, and is controlled by a diaphragm consisting of moving metal blades within the lens. A lens with a focal length of 100 mm (3.9 in) and an opening 50 mm (2.0 in) in diameter is said to have an aperture of 100/50 = f/2, but a 400 mm (15.7 in) lens with the same opening would have an aperture of 400/50 = f/8. Therefore, the size of the opening must increase proportionately with focal length in order for lenses of two different focal lengths to have the same aperture. This is why the large telephoto lenses used by sports and nature photographers are so long and wide. They must have both long focal lengths and wide openings to transmit enough light to properly expose the film.

The term f-stop refers to the fact that photographers have traditionally adjusted the aperture of their lenses by rotating a ring on the lens to choose among several pre-set apertures. Each pre-set aperture is marked by a sensible and audible click, or stop, hence the name f-stop. The pre-set apertures were chosen so that each stop halved or doubled the radius of the opening, using the same logic as pre-set shutter speeds, thus halving or doubling the amount of light passing through. The result was this progression of f-stops: f/1, f/1.4, f/2, f/2.8, f/4, f/5.6, f/8, f/11, f/16, f/22, and f/32. Although many modern lenses have continuously adjustable apertures, and some are electronically controlled with no aperture rings at all, the f-stop terminology and progression of f-stops marked on lenses persists. Physical constraints make it difficult to design lenses with large apertures, so the range of most lenses begins above f/1, typically in the range of f/2 or f/2.8. The additional difficulty of designing zoom lenses, especially if they are to be affordable to large numbers of people, sometimes motivates lens designers to use maximum apertures that change according to the focal length. An 18–70 mm (0.7–2.8 in) f/3.5-4.5 zoom lens would have a maximum aperture that ranges from f/3.5 at 18 mm (0.7 in) focal length to f/4.5 at 70 mm (2.8 in) focal length.

DEPTH OF FIELD

Depth of field refers to the range of distance from the lens, or depth, throughout which objects appear to be in focus. A lens can be focused on objects at only one distance, and objects closer to or farther away from the lens will be out of focus on the plane of the film. In the case of a point of light that is out of focus, the result is a fuzzy circle known as a circle of confusion. Depth of field is increased by decreasing the size of the circles of confusion in an image, which is accomplished by reducing the

aperture of the lens, until objects over a wide range of distances appear to be in focus to the human eye.

Although a small aperture reduces the sizes of the circles of confusion in an image, it also increases the relative importance of diffraction around the edges of aperture. As light passes through the movable metal blades that control aperture, some of it is scattered or diffracted. When the aperture is large, the effects of diffraction generally go unnoticed. As the aperture decreases, diffracted light becomes an increasingly large proportion of all the light passing through the aperture and image sharpness can decrease. Therefore, setting a lens to its smallest aperture will not generally produce the sharpest possible image. The sharpest images will generally be obtained by setting the lens to an aperture in the middle of its range.

For a specified aperture, lenses with long focal lengths will always have shallower depths of field than lenses with short focal lengths. This is because the longer lenses must have physically larger openings than the shorter lenses, even if the aperture (f-stop) is the same. A larger opening transmits more light, which in turn produces larger circles of confusion.

RECIPROCITY

Reciprocity is a mathematical relationship between shutter speed and lens aperture. If a photographer increases the light passing through the lens by opening the aperture one f-stop and then doubles the shutter speed (which will reduce the length of time the shutter is open by one-half), the amount of light reaching the film will not change. An aperture of f/4 and a shutter speed of 1/500 s, for example, will deliver the same amount of light as an aperture of f/5.6 and a shutter speed of 1/250 s. In other words, aperture and shutter speed share a reciprocal relationship and many different combinations of shutter speed and lens aperture will provide the same amount of light. The reciprocal relationship also extends to film speed. If film speed is doubled, either the shutter speed can be increased (producing a shorter exposure) or aperture can be decreased by the same factor without changing the amount of light that reaches the film.

In practice, there are some limitations to reciprocity. Photographs with very slow shutter speeds, for example minutes or hours instead of fractions of a second, can be appear too dark (underexposed) because the reciprocity relationship does not extend to such long exposures. This is known as reciprocity failure and can pose a problem for photographers working at night in situations where artificial lights cannot be used, for example when astronomers are attempting to take photos of the night sky using their very sensitive equipment. Film manufacturers publish tables that allow photographers to compensate for reciprocity failure in different kinds of film.

DIGITAL PHOTOGRAPHY

Virtually everything written in this article applies to digital photography as well as film photography. The primary difference is that a digital camera uses an electronic sensor instead of a piece of plastic film coated with silver compounds. In place of the film used in a conventional camera, a digital camera uses an electronic sensor. Two sensor types are commonly used: CCDs, or charged-coupled device sensors, and CMOSs, or complementary metal oxide semiconductor sensors. Both kinds are composed of rows and columns of photosites that convert light into an electronic signal. Each photosite is covered with a filter so that it is sensitive to only one of the three components of visible light (red, blue, or green). One widely used configuration, the Bayer array, consists of rows containing red and green filtered photosites alternating with rows containing green and blue photosites. When the image is being processed by the camera, values for the two missing colors are estimated using the mathematical technique of interpolation.

Two primary measures are used to characterize digital images: resolution and size. Resolution refers to the ability of a sensor to represent details, and is generally specified in terms of pixels per inch (ppi). Image size refers to the total number of pixels comprising an image, and is typically given in terms of megapixels. A pixel is the smallest possible discrete component of an image, typically a small square or dot, and one megapixel consists of one million pixels. As of early 2005, the best commercially available digital cameras had resolutions of approximately 20 megapixels and many professional quality digital cameras had resolutions of 5 or 6 megapixels.

Digital photographers can adjust the sensitivity of the sensor to light just as film photographers can use films with different ISO speeds. In digital cameras, however, there is done with a switch or button on the camera and the sensor is not physically removed. Although digital cameras commonly have ISO settings, they vary from manufacturer to manufacturer and do not follow the consistent ISO standard. Instead, they are an approximate gauge of the sensitivity. The digital equivalent of film grain is electronic noise, which can appear in images as visual static or randomly colored pixels, and is most often a problem using high digital ISO settings. The size of the sensor can also contribute to the amount of noise in a digital image, because the photosites on a small sensor are closer to each other than those on a larger sensor and can interfere with each other.

In order to capture action photos, photographers must use math to set shutter and film speeds properly. AP/WIDE WORLD PHOTOS. REPRODUCED BY PERMISSION.

A Brief History of Discovery and Development

The basic concept of using a device to project an image onto a flat surface dates from the camera obscura of ancient times, in which light passed through a small hole that focused the image and projected it in a darkened room. The modern day descendent of the camera obscura is the pinhole camera, which uses a hole without a lens to project an image onto a piece of photographic film. The quality of camera obscura images increased as lenses were developed in the sixteenth century. Still, there as no way to preserve the image except by drawing or painting on the projection screen. The discovery of photosensitive chemicals in the nineteenth century was a major step forward because it allowed images to be preserved without drawing or painting, and many different techniques were invented for creating photographs on paper, glass, and metal sheets. In 1861, Scottish physicist James Clerk-Maxwell invented a system of color photography using black and white images taken through red, green, and blue filters and then combined. George Eastman started his photographic company in 1880, and the first Kodak camera was introduced in 1888. This surge in technology gave rise to an explosion in the technical, journalistic, and artistic use of photography as mechanical cameras and lenses were continually refined throughout the first half of the twentieth century. The advent of computer-aided design in the 1960s and 1970s represented another major step forward, allowing much more sophisticated camera and lens designs, and auto-focus and auto-exposure cameras arrived on the scene shortly thereafter.

Real-life Applications

SPORTS AND WILDLIFE PHOTOGRAPHY

Sports and wildlife photographers often share the same goals. They want to produce photographs of fast moving subjects from a distance. Therefore, they prefer long telephoto lenses with large maximum apertures. By

Key Terms

Aperture, lens: The size of the opening through which light passes in a photographic lens.

Reciprocity: The mathematical reciprocal relationship between shutter speed and aperture, which states that there are many combinations of lens aperture and shutter speed that will supply the same amount of light to the film or digital sensor in a camera.

virtue of reciprocity, these large aperture lenses can be used with higher shutter speeds that freeze action, whether it be a gazelle or a linebacker. Long lenses with large maximum apertures also add an artistic element, helping to blur the background and focus the viewer's eyes on the subject of the photograph. For the same reason, portrait photographers will often used moderately long telephoto lenses that set their subjects apart from the background. Photographers describe the aesthetic quality of the blurred areas with the Japanese word *bokeh*, and a lens that produces pleasingly out-of-focus areas is said to have good *bokeh*.

DIGITAL IMAGE PROCESSING

The ability to create high-resolution digital images, either using a digital camera or by scanning a film negative or transparency, allows photographers to adjust the details of their photographs without entering a darkroom. Each pixel contains a red, green, and blue value that can be brightened or darkened. The overall range of tones, known as contrast, can also be easily adjusted and unwanted tints can be removed. A photographer, for example, can remove the cool bluish cast in shadowy light by adding more red and green to the image. Images can also be sharpened to some degree, although it is impossible to sharpen an image that is truly out of focus. This is done using a technique called unsharp masking, which derives its name from a technique developed by astrophotographers using film many years ago. In order to sharpen a slightly fuzzy image, the photographer would make a deliberately blurred copy of the film negative. The two images would be carefully aligned and a sharpened print made.

PHOTOMICROGRAPHY

Photomicrography uses an optical microscope, rather than a traditional lens, to produce photographs of objects such as microorganisms and mineral grains. In the case of geological photomicrography, small slices of rock are glued to microsope slides and then ground down to a thickness of 30 microns (0.001 in). The slice of rock is nearly transparent at that thickness, allowing it to be examined under the microscope. Digital image processing techniques can also be applied to photomicrographs in order to enhance edges or increase the visibility of subtle details.

Potential Applications

Computer designed lenses and cameras, both film and digital, continue to increase in sophistication each year. Current commercial activity emphasizes the development of improved digital sensors with increased resolution and decreased noise, vibration resistant camera bodies and lenses that compensate for the photographers moving hands, and zoom lenses that cover focal length ranges from wide angle to telephoto.

Where to Learn More

Books

Enfield, Jill. *Photo-Imaging: A Complete Guide to Alternative Processes.* New York: Amphoto, 2002.

Jacobson, Ralph, Sidney Ray, G.G. Attridge, and Norman Axford. *Manual of Photography: Photographic and Digital Imaging,* 9th ed. New York: Bantam, 1998.

Web sites

Greenspun, Philip. "History of Photography Timeline." 2005. <http://www.photo.net/history/timeline> (February 15, 2005).

Kodak. "Photography in Your Science Fair Project: Photomicrography." No date. <http://www.kodak.com/global/en/service/scienceFair/photomicrography.shtml> (February 15, 2005).

Plots and Diagrams

Overview

The use of plots and diagrams is an integral part of everyday life. Plots and diagrams can be found in many applications in scientific study and in real life.

Effective graphs can significantly increase a reader's understanding of complex data sets. The basis of scientific procedure is data collection. Scientists are required to examine and analyze the data they collect. The most efficient way to do this is graphically. A graph is a visual representation of two variables relative to each other. Graphs are one- or two-dimensional figures. Three-dimensional graphs also exist, however, these are often more complex and more difficult to understand than basic two-dimensional graphs. A graph usually has two axes, the x-axis and the y-axis. There is also an origin, which is the point (0, 0). This is where the two axes cross each other. Each point on a scatter graph is represented by a pair of coordinates. These are written in the form (x, y). The number x represents how far along the x-axis the point is, and y represents how far along the y-axis the point is. If a point lies on the y-axis, its co-ordinates would be (0, y), because it is at the 0 point along the y-axis (remember the axes cross each other at 0). Accordingly, if a point is on the x-axis, then its co-ordinates are (x, 0).

PROPERTIES OF GRAPHS

A graph should have at least a title and a scale that is numbered in specific and constant intervals and labeled. This allows the reader to know what the graph is about and what the graph is measuring or showing. The more information that is included on the graph, the easier it is to understand and interpret the data it shows. However, too much information must not be included, as the graph may become cluttered. Some graphs require a legend or key. This helps the reader understand different shading and colors that have been used. A legend is useful if the graph becomes too cluttered with all the labeling. The purpose of a graph is to provide clear, concise information. This is difficult to accomplish if there are large numbers of labels covering the data.

DIAGRAMS

Diagrams can come in many shapes and forms, depending on the application for which they are being used. Most graphs are about numbers; in other words, they are number oriented. However, with diagrams this need not be the case. Some diagrams do present quantitative (number oriented) data, but most diagrams present qualitative (non-numerical) data. They are widely used in

both science and everyday life. The type of diagram directly depends on the subject data. Diagrams are usually pictures. Around or on this picture is usually written extra information. This information could be providing details about the diagram, such as a diagram of the body, or the diagram could be there for easier comprehension of details, such as a weather map.

Fundamental Mathematical Concepts and Terms

STEM AND LEAF PLOTS

Stem and leaf plots are similar to histograms (vertical graphs with touching bars) in the way they represent information. However, they usually contain a little more information. Stem and leaf plots show the distribution (or the shape of the data) as well as individual data. These types of plots are useful in organizing large groups of data. In a set of data containing numbers from 1 to 100, the digits in the largest place, the tens, are referred to as the stem. The digits in the smallest place, the units or ones, are referred to as the leaf. When there is a large amount of data, sometimes the stem needs to be represented twice. The first time it is associated with the leaves 0 to 4, and the second time it is associated with the leaves 5 to 9. If a stem is shown five times, then similar rules apply as when it is represented twice. The first stem is associated with 0 to 1, the second with 2 to 3, and so on. This is to make the plot easier to read.

BOX PLOT

A box plot (also known as a box and whisker plot) is a diagram of the measure of spread. It is a graph of the 5-number summary. Data can be divided into four even sections called quartiles. The number of values in each quartile is the same. The middle number is called the median. The value between the median and the minimum value is the first quartile and the value between the median and the maximum value is the third quartile. The 5-number summary is the minimum, the first quartile, the median, the third quartile and the maximum. The inter-quartile range is the distance from quartile 1 to quartile 3. A quartile is 25% of the numbers of the entire set of data. A box plot shows the spread of a set of values. This is an important factor in some statistical analyses.

SCATTER GRAPH

Scientists most often utilize scatter graphs. They are useful for fast and easy analysis of data. These types of graphs are usually a series of points on a grid. Each of the axes is used to represent a value data. The value of the variable along the y-axis (the vertical axis) is dependent on the value of the variable along the x-axis, which is the independent variable.

Scatter graphs are usually used to determine a relationship between two variables. Once two sets of data have been plotted against each other (such as distance against time), a line of best fit can be drawn through the points to determine whether there is a relationship between the two variables. Scatter graphs are most commonly used for scientific purposes. This is because they do not negate individual data. Every single piece of data is included in a scatter graph. However, scatter graphs can also show two sets of data that had the same variables measured, but one was changed.

Three mathematical concepts that are unique and integral to scatter graphs are the line of best fit, the correlation coefficient, and the coefficient of determination. These three tools are important in helping scientists analyze the data that they gather. In real-life applications, the interpretation and understanding of data is the most important part of scientific process. Without interpretation, and thus tools of interpretation, data would just be a meaningless set of numbers.

A line of best fit, also known as a line of regression, is a line that is drawn to represent the trend of the data. A regression line always exists, whether there is correlation, a relationship between two variables, or not. The easiest way to draw this line is to draw a straight line through as many points of data as possible. However, this is usually impossible, especially when scientific errors are taken into account. Then the best method to draw this line is to have an equal number of points above the line and below the line. This averages out the line. There are complicated methods of determining the exact line of best fit that involve long and laborious calculations. A line of best fit is where the vertical deviations (the up or down distances) from the observed point (the ones determined experimentally) and the calculated points (the ones taken from the regression line) are as small as possible. In other words, the line of best fit is a refined line of regression, although the two terms are usually used to represent the same line. It would take a long time to determine the line of best fit if drawing the line of regression by hand. Computer programs for data analysis exist now that can compute and draw the line of best fit automatically. The computer does all the calculations much faster than a person would be able to do it.

The correlation coefficient is an important concept to understand when interpreting graphs and their lines of best fit. The correlation coefficient is a way to measure

how close the points are to a regression line. The correlation coefficient is commonly known as r and lies between -1 and 1. When $r = \pm 1$, then there is perfect correlation between the two variables and all the points lie on the line. Then $r = 0$, there is no correlation between two variables and they are all independent of each other and the line of regression. A correlation coefficient between 0.0 and ± 0.3 is considered a weak, a correlation coefficient between ± 0.3 and ± 0.7 is considered a moderate, and a correlation coefficient between ± 0.7 and ± 1.0 is considered a high. Mathematically, the correlation coefficient is the sum of the squares of the individual errors, which are the vertical deviations, to measure how well a function, usually the line of regression, predicts y from x.

The coefficient of determination, R^2, is another measure of how well two variables are related and how well a regression line fits to the set of points. R^2 describes how much of the deviance in the y values can be explained by the fact that they are related to an x value. In simple linear regression, R^2 is the square of the correlation coefficient, in other words, $R^2 = r^2$. Both of these coefficients can help people determine whether data is credible or not, especially in a scientific context. Usually a scientist will have a thesis or aim that he or she wants to prove or disprove. Here the correlation between height and arm span will be used and the aim is to show that they are related to each other. The scientist will take an ample amount of data and then analyze this data, probably using a scatter graph. If the correlation coefficient or R^2 value is below standard to prove the aim correct, then the scientist may have to revise the data or gather more data. This process, especially the R^2 value, is integral to the process of scientific information, especially if a scientist is looking to present credible data.

AREA CHART

A variation of the line chart is an area chart. Line charts look like various line graphs together with the sections between them colored in. They are used where there is one independent series and several dependent series. The independent series together have a constant sum.

PIE GRAPH

Another type of graph is a pie graph. These graphs are aptly named, as they have a circular shape and sections are cut separated by a line making the whole graph look like an unevenly cut pie. The idea is effective because it takes advantage of the everyday principles people use when, say, they are cutting a cake into portions. This makes the pie chart something people can relate to and thus more easily understand. Although these graphs are not often seen, they are the most useful in expressing discrete data in specific categories. They are used to show how one piece fits into the entirety; in other words, pie graphs are used when the values have a constant sum, such as a population or when using percentages.

Pie graphs are best utilized when there is significant variation between the portions. In other words, having five equal areas is quite useless, unless that is the point being made. Pie graphs often have the sections labeled directly on the diagram instead of having a separate table informing the reader of which section is which. However, a pie graph does have its limitations. The number of categories (or portions of the pie) needs to be small or the graph may become cluttered. Generally, the number of categories should be between 3 and 10, although this may vary slightly.

BAR GRAPHS

Bar graphs are another popular style of graph. Bar graphs are a versatile type of graph and can show many different types of data. A bar graph, also called a column graph, is easy to recognize because several long or short rectangles represent data categories. These rectangular bars can be vertically or horizontally orientated. A bar graph has the bars orientated horizontally and a column graph has them orientated vertically. Sometimes this distinction is not made, however, it is important to know the difference. On a bar graph, the bars are usually the same width. They are used as a comparative type of graph and usually compare several things, people, objects, cities or departments, units or entities, in a data series. On a column graph, these categories are along the horizontal axis whereas on a bar graph they are on the vertical axis.

FISHBONE DIAGRAM

Fishbone diagrams have a strange name that resembles the style of graph. Fishbone graphs are primarily used in problem solving, especially in quality improvement programs in business. The problem is written inside a box at the head of the fish. This provides the aim for the diagram. The words and ideas that extend from the backbone are the possible causes of the problem. They allow people to organize thoughts about problems and what may be causing them. It is a systematic way to organize and analyze data that is related to solving a quality problem. Fishbone diagrams may also be called cause-and-effect diagrams or Ishikawa diagrams.

POLAR CHART

There are several specialized types of graphs that have been developed to deal with unusual and different

types of data. Even though scatter graphs, pie charts, bar charts, and line graphs deal with a wide variety of data, sometimes even they cannot handle specific data. One such graph that has been developed is the polar chart. A polar chart is used with discrete data where each point has a direction value from a source, in other words, a direction, usually expressed in degrees. The data also has a quantity, or a specific distance it is away from the source. Essentially, this graph represents polar data. Polar graphs are used in the study of polar equations and also in vector studies.

TRIANGULAR GRAPH

Another specific type of graph is a triangular graph. These graphs are most commonly used in geographic applications. They are used to plot discrete data in which each point has three values. These values have a constant sum that is usually expressed as a percentage. It is triangular in shape, hence its name, the triangular graph.

THREE-DIMENSIONAL GRAPH

A more complex type of graph is a three-dimensional graph. These are used when three interdependent variables need to be plotted. If data are grouped, then a three-dimensional column graph can be used. This allows the reader to associate certain things relative to others in one graph instead of having to use many different graphs. When data is displayed in this manner, it may initially look quite complex. However, with further understanding of what the graph represents, it is a suitably useful way of displaying data. If data are from a continuous distribution then a surface plot is used. This is another type of three-dimensional plot. These are more difficult to interpret, as the data is continuous. Three-dimensional graphs can be shown as a continuous surface or as a series of contour lines.

A Brief History of Discovery and Development

The origins of plots and diagrams date to prehistoric ages when people made cave drawings. Although these may not be the sophisticated plots and diagrams that exist in the twenty-first century, they were, nonetheless, diagrams. The earliest map was a tenth century map of China. Diagrams of planets and planetary motions were some of the earliest, more complex diagrams that existed. Graphs made their first appearance around 1770, and became accepted and widely used around 1820. In 1795, graphical scales were used to help convert old measurements to the new, metric measurements. The French mathematician Johann Heinrich Lambert (1728–1777) used graphs extensively in the eighteenth century. He was one of the only scientists of the time to do so. He applied many of the principles now applied to graphs, such as a line of best fit. From this time, graphs and diagrams developed to aid people in determining angles, analyzing data, and providing information. Bar graphs emerged for data that could not be sorted. General x-y graphs did not appear in publications until the twentieth century. From simple things, such as pictorial instructions, to complex graphs, everyday life has been greatly affected by plots and diagrams.

Stem and leaf display

3	2337
2	001112223889
1	2244456888899
0	69

Figure 1: A stem and leaf plot display.

Real-life Applications

STEM AND LEAF PLOTS

Stem and leaf plots can be used for series of scores on sports teams, series of temperatures or rainfall in a month, or series of classroom test scores. In a stem and leaf plot, data is arranged in place value. (See Figure 1.)

BOX PLOT

Box plots are usually drawn as composite box plots. (See Figure 2.) Two different box plots displaying, for example, the heights of boys in a class and the heights of girls in a class, can provide more statistically useful data when compared than a single box plot. It is a simple and clear graphical representation of information that may be difficult to decipher as just a series of numbers on a page. The two graphs together allow the reader to easily interpret the ranges of girls' height with respect to the boys, and vice versa. Box plots also show whether there is a data point that is an outlier, that is, it does not fit within the specified set.

SCATTER GRAPH

Scatter graphs are used to plot much of the experimental data that scientists collect. For example, if a scientist

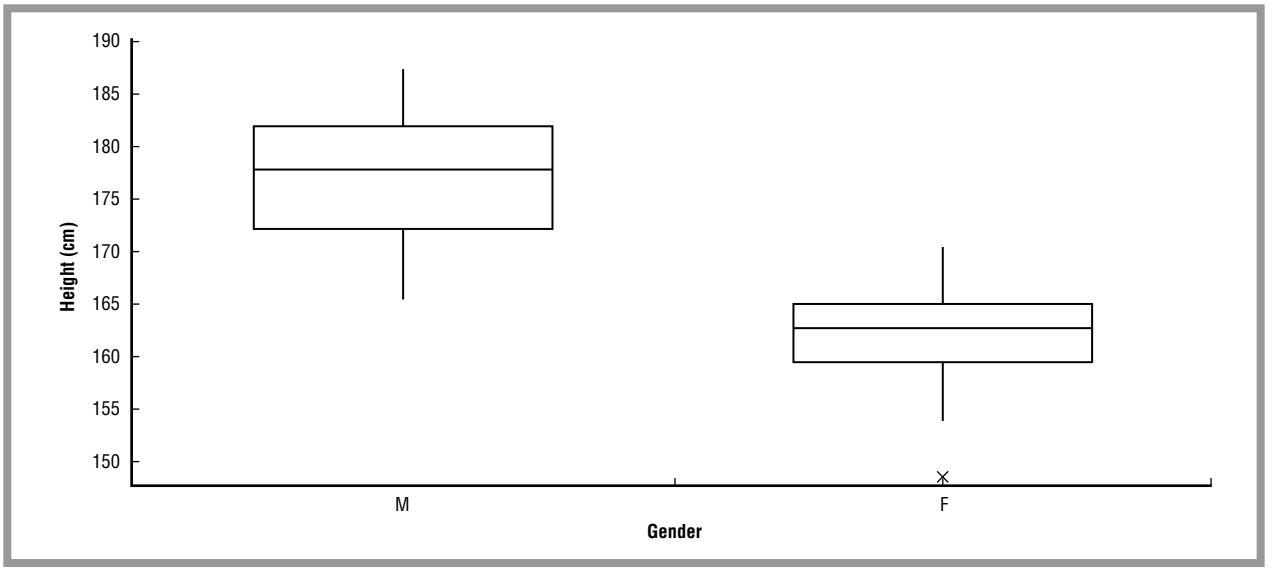


Figure 2: A composite box plot.

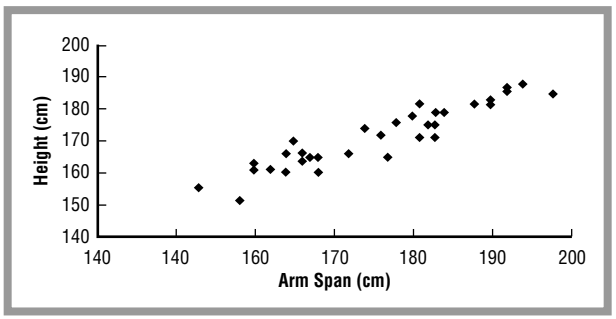


Figure 3: A scatter plot on a scatter graph.

measured the distance a car, traveling at a constant speed, traveled over a period of time, a scatter graph would measure the time and the speed. In this case, the time is the independent variable because the distance the car travels depends on the amount of time the car is traveling. When comparing two sets of data, a scientist could compare the braking distances of cars traveling at different speeds. The speed the car is traveling at would be the independent variable, and the braking distance would be the dependent variable. Different brands and types of cars can be shown on a scatter graph to compare and contrast them to each other. This can help a person wishing to purchase a car to determine which car they want to buy if safety is one of their primary buying criteria.

A scatter graph can also be used to determine if there is a correlation between a person's arm span and their height. The initial scatter graph would show the grouping of the data and can show whether there is evidence of a trend or not. (See Figure 3.)

Once a set of data is plotted, a line of best fit can be drawn to show whether the relationship between two variables is worth investigating. An example would be a scatter graph illustrating the growth of a rabbit population. The population was counted at various intervals and the data was plotted on the graph. A line of best fit was drawn and the correlation coefficient was determined. The use of scatter graphs can help scientists determine important statistical data. People can then use this information when they are studying the growth of populations for assignment or more in depth studies. These graphs help people to understand the way things work without having to be scientists. (See Figure 4.)

LINE GRAPH

Line graphs represent changes in numbers over time. It is one of the most widely used types of graphs in real life. Stock market graphs are an example of line graphs. They provide stockbrokers and the general public with long-term information about a particular stock. It is easiest to read how much a particular stock has gone up in one day. However, to determine whether the stock would be a good investment in the long-term, it is best to look back at years of data. However, poring over page after page of numbers is difficult and laborious, so this data is best represented in a scatter-plot that has each point joined together. This enables people to tell whether its value has fallen or risen over a period of time, all with just

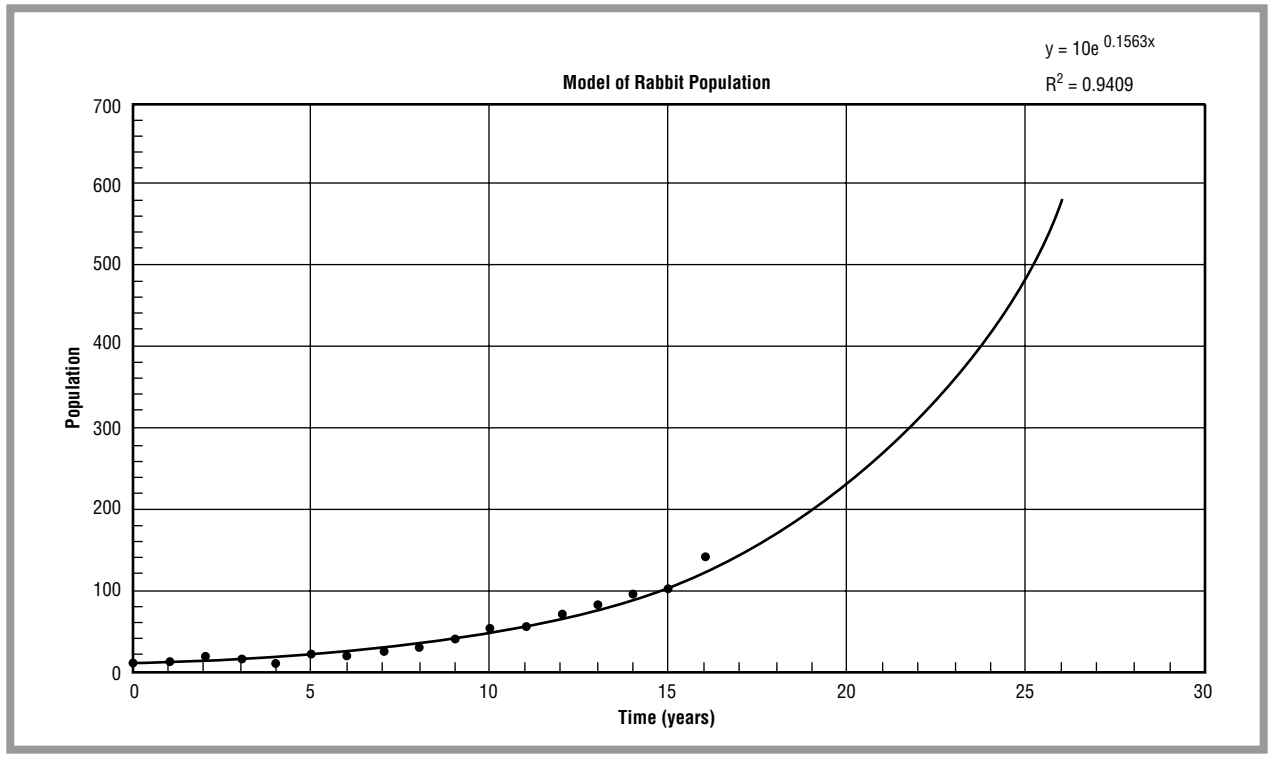


Figure 4: A scatter graph with best line averaging points and transitions between data point that illustrates the growth of a rabbit population.

one look at a single image. Stock market graphs are found in newspapers and can be seen on television. Although these may be difficult to understand to the novice, for a person who wants the information and who understands the basic principles, they are easy to decipher and quick to understand.

Another example of a line graph is to show the number of people who attend sports matches over a period of time. Time would be illustrated on the x-axis, which means that the number of people who attend is dependent on the time. The graph would show how many people attended matches each year. It would show how the trend the amount of people who attended matches increased or decreased. It would also show specific slumps and rises in attendance numbers. These graphs can have future applications in determining the reasons for lack of attendance or for high attendance. This would require further research, such as looking at world events that may have caused particular slumps or high points.

Figure 5 depicts a line graph as a composite line graph. This means that it shows different types of data that have the same variables measured and are comparable. In this case, it is the growth of a normal mouse relative to that of mutant mice.

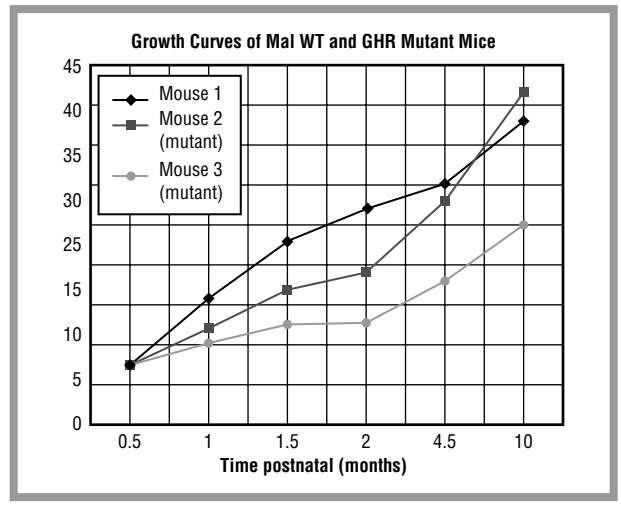


Figure 5: A line graph as a composite line graph.

A special type of line graph is a run chart. Run charts show the sequential measurements of a process over a specific period of time. One example is how many cookies are produced from a batch of dough that is presumably the same size each time. These types of graphs usually have limits, depending on what is being measured.

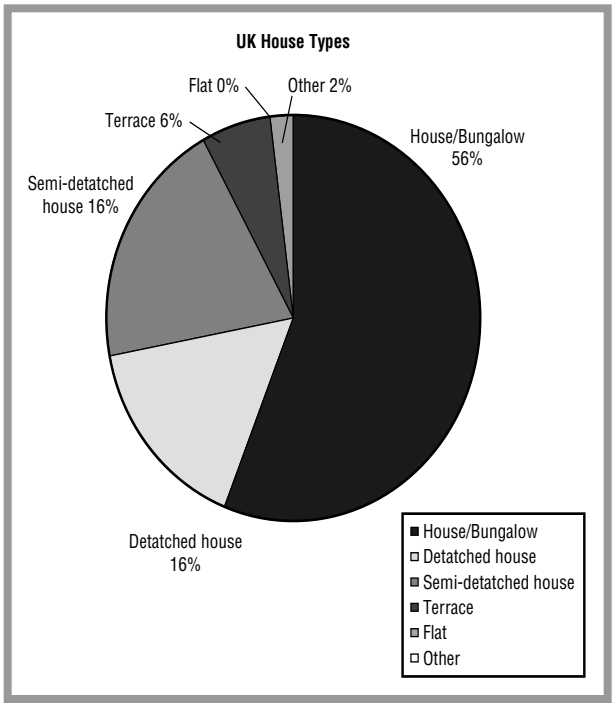


Figure 6: A pie graph can also be used to compare the different types of housing in which people live.

Another example could illustrate patient waiting time in a dentist's or doctor's office.

PIE GRAPH

Pie graphs can be used for such purposes as illustrating the different cultural backgrounds of people in a country. The categories in the pie chart are often linked, such as cultural backgrounds, as in the example of Figure 6 given. Pie graphs provide a more overall view of data and can be difficult to read if the data is not written on or near each pie section. However, they are useful in determining general trends or providing rough estimates. This is because pie charts are quite graphic and tend to be colorful, to make distinguishing between sections easy. They are useful in representing percentages of stock in a store of percentages of a population that have particular diseases. As shown in Figure 6, a pie graph can also be used to compare the different types of housing in which people live.

BAR GRAPHS

Bar graphs are used in a similar fashion to pie graphs in that they can express discrete data in specific categories with great efficiency. Bar graphs are more useful than pie graphs when exact numerical data is more important rather than a general overview. For example, the number of people who answered a particular multiple-choice question would be best represented in a bar graph. This would allow students to easily compare their own scores with other students' scores and also determine whether they were in a majority or not. A simple bar graph, as shown in Figure 7, quickly shows the frequency of specific heights of people which can then be used to compare people's heights.

Another example of how a bar graph can be used is depicted in Figure 8, a graph that compares technology access.

As shown in Figure 9, another method to depict data joins the successive midpoints of each of the bars to make another type of graph, a frequency histogram. Histograms provide information on the distribution of the data, a concept that relates more to statistics than to graphs and diagrams.

FISHBONE DIAGRAM

If a person in a business has a problem, such as low production, a fishbone diagram can be applied to determine the cause of the problem. The problem is written in the head of the fish. From the backbone of the fish are bones, each with a specific topic, such as psychological factors, company pressures, physical pressures, and home problems, for a set example. From each of these are the different contributing factors within those topics. This visual representation can help both the person and the company locate and identify the problem and thus start the process of solving the problem.

TRIANGULAR GRAPH

Each side of the triangle is labeled with a different soil type, for example. Along that line are marked the various percentages, from 0% to 100%. A point marked inside the triangle denotes a specific soil type. Its composition can be determined by drawing a line from it to each edge of the triangle. Where the line lands perpendicular to the side of the triangle is the percentage of that compound that can be found in that soil. This process can also be used in reverse by knowing the percentages of the composition and then determining the specific soil type from those percentages. Its most common use is the three-way split of sand, silt, and clay in soil and sediment samples. This type of diagram allows scientists to determine the type of soil that it is, and thus postulate where the region may have been developed (since land masses have moved due to continental drift). It can also help them determine what uses the soil may have. This means that this type of graph can help a person determine the

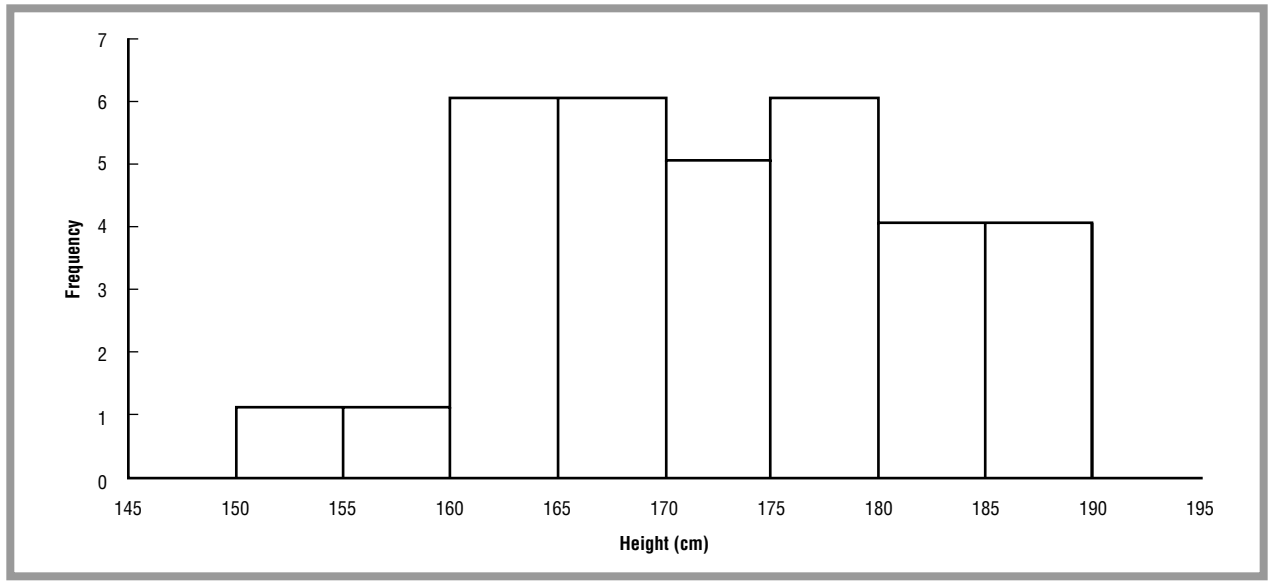


Figure 7: A bar graph.

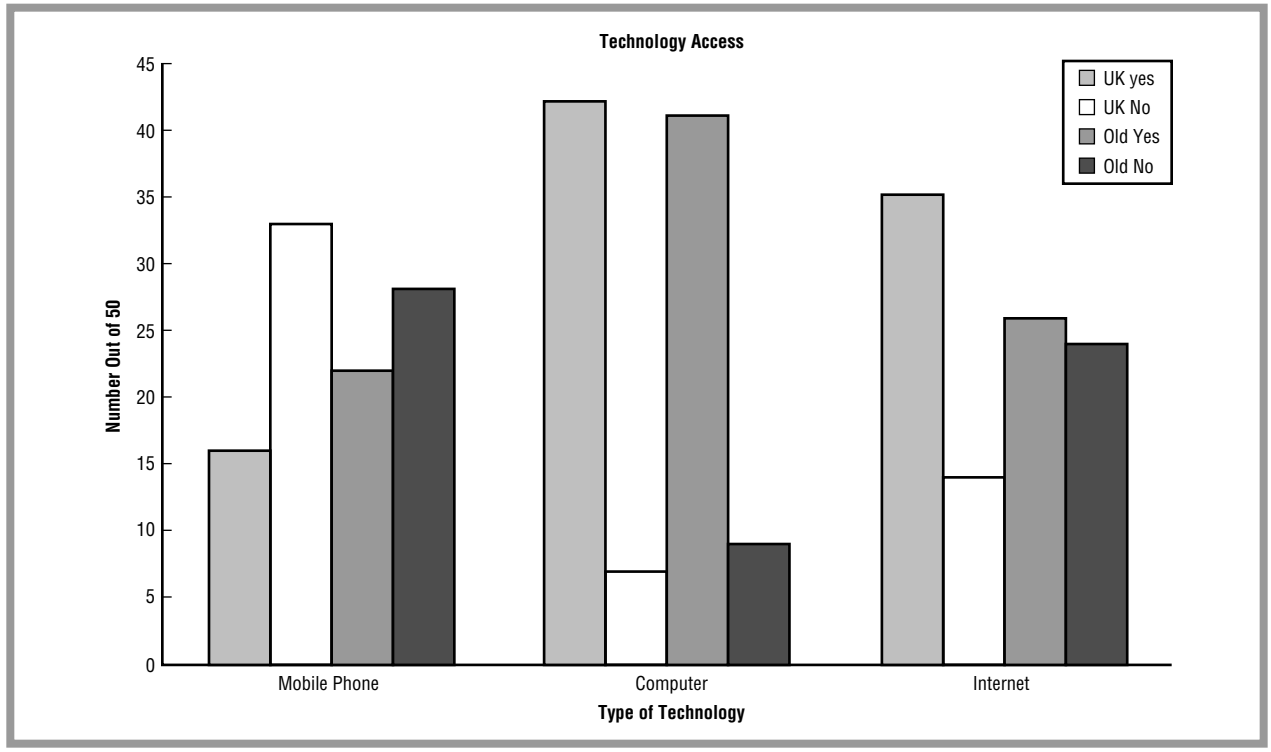


Figure 8: A bar graph comparing technology access.

type of soil they need for their garden to make sure their plants have the correct soil. However, this type of graph can also be used where three factors are needed to represent a whole and each is of a specific percentage.

FLOW CHART

An important type of diagram is a flow chart. Flow charts are useful because they show sequential steps in a process. They are good reference points for people learning

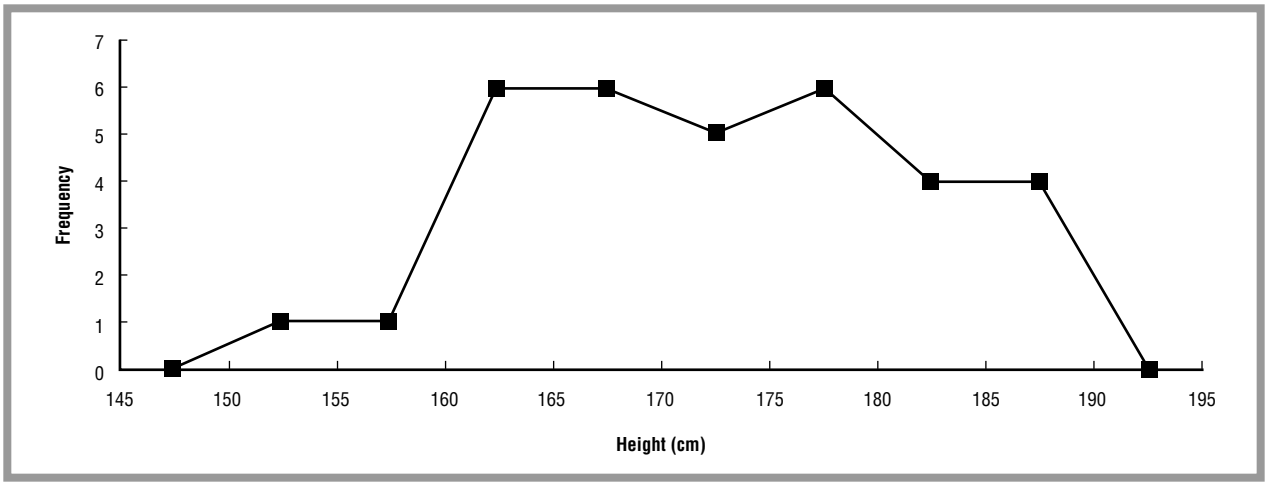


Figure 9: A frequency histogram.

a new skill or job. They are used to represent a group of connected components in a series where there may be more than one pathway through them. Occasionally, flow charts have a web-like structure. Quantitative aspects of the data flow can be distinguished by varying the line style or thickness between the components, and labeling each of the components thoroughly.

A flow chart has many applications. One of the most important applications is in computing where flow charts are used to represent the internal logical organization of computer programs. Examples of these charts can be seen on a personal home computer, or PC. When navigating through folders to find a file, one of a multitude of possible paths is followed. There are also methods for determining the number of different pathways, however, these are complex and are not discussed here. Flow charts are also used in business organization. One example is processing a sales transaction, from the flow of information and goods beginning at the receipt of the order, to the shipping and the invoicing of the customer.

New computer systems are larger, can process more data, and can accomplish this at a greater speed than older models. This means that more information can be stored on them and thus, more information needs to be accessed and processed at a comparative or faster speed. Alternative diagrammatic structures have been developed to aid system analysts in designing information processing systems that have quite complex internal information flows. The tools allow the analysts to model the relationships amongst the components of a system. These relationships can become quite complex, especially with new computer systems. These tools and structures are all encompassed under the general heading of relationship diagrams.

Representations of the World Wide Web, more commonly referred to as the Internet, are usually drawn as flow charts. This is due to the large amounts of interconnected data that exist on the Internet. The concept of virtual realities is based upon this interconnectivity. Virtual realities are becoming more and more a part of everyday life for people. Computer and arcade games use virtual reality concepts. Virtual realities have the ability to allow the user to experience things they may not be able to experience in real life, such as river rafting, snow skiing, or water skiing. Behind the graphical images of virtual realities are flow charts not unlike those that have been described. Each choice in a virtual reality game, for example, is a path that is followed though an interconnected web, in other words, a flow chart. One decision is simply just that, one decision, meaning that each time the game is played and one decision is made differently, then the entire game experience may be different. A flow chart is a diagram that can be used to map the progress of person's experience in a virtual reality game, or map Internet surfing or web structure as depicted in Figure 10.

This flow chart is an example of a Web page. Some flow charts illustrate a single flow. However, others can be drawn to represent the relationships between each component, in this case a web page, and thus, the various paths that can be followed through a web page via hyperlinks. Flow charts represent these complex data paths in simple, understandable ways.

TREE DIAGRAM

A tree diagram shows the relationships amongst main concepts and contributing concepts. Tree diagrams

usually begin with a main concept or idea, such as an essay topic. From this main concept, several sub ideas branch off (hence the name tree diagram). From these, more information branches off until all the desired information is included. Tree diagrams can be vertical, which means they have the initial concept at the top, or horizontal, where the initial concept is on the right hand side.

A tree diagram can be used in statistics to determine the probability of throwing two heads and two tails when flipping a coin. A family tree is another example of a tree diagram. Some family trees are oriented to have the last person born at the bottom; this represents a reversed vertical tree diagram. However, the orientation does not detract from the purpose or meaning.

ORGANIZATION CHARTS

Organization charts are slightly different than flow charts. Flow charts have a web-like structure, and organization charts tend to be less interconnected. Organizational charts usually have a hierarchal structure. This means that there is a leader at the top, and subordinates are illustrated below. Usually, as the chart progresses down the hierarchy, each level contains more entries.

The primary place where organization charts are used is in organizations or business firms. These present the hierarchal structure vividly. They are used best when there are more than three people in an organization. Organization charts can also be used to map the evolution of objects such as refrigerators, microwaves, and computers. They will show the earliest model at the top of the chart, and then year-by-year how new models evolved from the older models. Organizational charts are easy to understand and are much simpler to read than reading a paragraph or essay about the history of the appliance.

GANTT CHARTS

A Gantt chart is a time-management and product-management tool in a business that helps organize an individual or a group. It is a visual presentation of parts of a project and how they all relate to one another. Gantt charts show the progression of a project or a specific task to be completed. They use a timeline along the top and a list of tasks or people down the side. Different tasks may also be represented in different colors. It can help employees to determine who is relying on them and on whom they are relying to have work completed. Gantt charts also provide a visualization of project deadlines. They are most useful for tracking work, scheduling work, and planning work for a project, especially when there are several people involved.

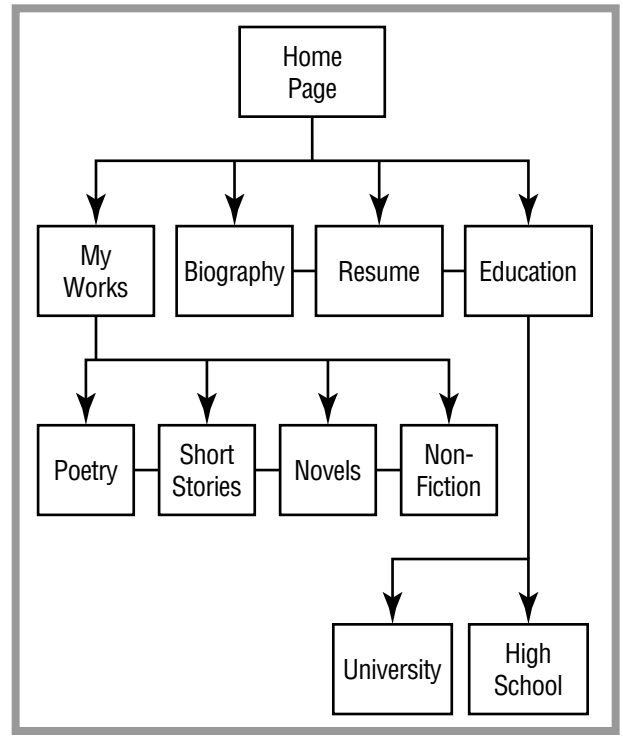


Figure 10: A flow chart.

MAPS

Maps are a type of diagram. They represent a certain area with a picture by showing the placement of one thing relative to another. In the case of road maps, they are usually labeled and contain important information about how to travel to a destination or where that destination is located. A map is a complex diagram consisting of various pieces of information portrayed in different colors and symbols that are explained in a legend.

A map is easier to read and understand than a complex set of instructions, especially for long distances. It provides an easy to view comparison of places relative to one another. Maps also provide specific information about amenities in a local area. They can show where parks are located, where libraries are, or where shopping centers are. World maps show the different continents and how they are arranged. Some maps show the world's ocean currents, others show the tectonic plates and the direction in which they are moving. Some maps are topographical, meaning that they have lines that represent different heights and they also show most, if not all, the specific details of the terrain.

There are many potential uses for maps. Every time a new estate or town is built, a new map has to be drawn. There is always going to be a need for maps. Now maps are being used in global positioning systems (GPS) to

help determine specific locations. These are especially useful when a person cannot be found because they are buried under an avalanche or in a collapsed building. The maps incorporated into GPS help rescuers find these people.

WEATHER MAPS

Other commonly used maps are weather maps, which can illustrate areas worldwide, country-wide, and region-wide, along with local areas. Weather maps are a simple and easy to understand way to present data about the weather. Instead of having the wind speed and direction, the temperature and weather forecast for every city, and the isobars, the weather map presents this information in a single screen on the television or a single image in the newspaper. To look at a weather map is easier than to read all of the information.

Many types of maps are different from geographical maps. For example, a calendar is a map, and thus, by definition, a type of diagram. It shows the days in a month or year relative to one another and allows people to keep track of time. A diagram of the moon phases is also a diagram and can be thought of as a map. A map is an object that simply shows one thing relative to another. Therefore, the term map is quite versatile and includes many diagrams.

BODY DIAGRAM

Diagrams often seen in real life include those of the human body. For centuries, physicians attempted to treat patients without proper knowledge of human physiology. This led to physicians being feared by many people, and patients benefiting little from physicians. However, this was in the Dark Ages when medicine was a more superstitious than scientific practice. In the twenty-first century, physicians have many different options for treating patients, including x-ray images to determine where bones are broken. X rays are a type of diagram.

STREET SIGNS

Some of the most useful diagrams occur in day-to-day life. One of the most common diagrams seen by people every day is the street sign. Street signs denote instructions for traffic and pedestrians. Without street signs, traffic would be disorganized and safety would be compromised.

CIRCUIT DIAGRAM

Another vital diagram is an electrical circuit diagram. An electrician has a diagram of all of the circuits in a person's home. This allows the electrician to complete work. The diagram enables him to determine which circuit he must enter to fix an appliance or lighting fixture. It saves him from having to switch off all power to the house, and may also save his life. These diagrams need to be accurate and up to date; otherwise people's lives may be at risk.

Circuit diagrams are most often seen in computer-related work. Although most students have only experienced circuits that involve minimal components, such as a 9-volt battery and a small light bulb, circuits can become quite complex and difficult to decipher. Diagrams are infinitely useful in determining both the function and size of a circuit. Electrical circuits are everywhere, in computer, cars, refrigerators, washing machines, dishwashers, mobile phones, Discmans, and many more household items. The circuits found in everyday objects are formed from circuit diagrams. These diagrams allow manufacturers to manipulate the size and shape of circuits before they actually make them. This becomes more efficient with the use of circuit diagram drawing programs on computers, which save time and resources in the production stages. On a circuit diagram, different symbols denote different circuit components. Two vertical lines, one shorter than the other, denote a battery or cell, whereas two equal length lines denote a capacitor. A circle with a cross in it denotes a light bulb. There are thousands of components in electrical circuits. Each of them has a different symbol. A circuit diagram arranges these symbols in an order, so that they perform a particular function.

OTHER DIAGRAMS

Lighting designers use diagrams of lights to illuminate a Broadway spectacular. These diagrams show the type of lights and their placement. A lighting bar diagram allows a lighting designer to design the lighting for a show without the trouble of rigging and re-rigging the lights each time they want to make a change. This diagram helps the designer to work out where the lights should be initially placed. Through experience, the designer would be able to produce an effective design without having to rig any lights.

There also exist diagrams that are instructions on the use of a particular item. Instructional diagrams may be included in a manual for the operation of unfamiliar technology such as digital cameras. These manuals contain diagrams that provide important information on the functions of specific buttons. They also show what a button might look like and its placement on the camera in the diagrammatic instruction, so that the reader is easily able to discern the function and the button required for it. There are diagrams that explain the symbols used on a particular compact disc player. Diagrammed instructions also help when assembling consumer products such as

furniture, toys, or bicycles. Putting together a bookcase, for example, would be difficult without a diagram showing where to attach the shelves.

Where to Learn More

Books

David, C., and S. Moore. *The Basic Practice of Statistics*. New York: W.H. Freeman and Company, 1995.

Weaver, Marcia. *Visual Literacy: How to Read and Use Information in Graphical Form*. New York, NY: Learning Express, LLC., 1999.

Web sites

Egger, Anne C. "Visualizing Scientific Data: An essential component of research." Visionlearning Vol. SCI-2 (1), 2004. <http://www.visionlearning.com/library/module_viewer.php?mid=10> (Mar. 30, 2005).

"Graphs." University of Rhode Island. <http://www.uri.edu/artsci/ecn/mead/306a/Overviews/overview.Graph.html> (Mar. 30, 2005).

Powers

Overview

The term "powers" is used to describe the result of repeatedly multiplying a number by itself. It is represented by: $a \times a \times a \times a \ldots a$, where 'a' is a variable, '×' shows the process of multiplication, and '. . .' shows that the process can be repeated some number of times. For example, nine (3×3), twenty-seven ($3 \times 3 \times 3$), and eighty-one ($3 \times 3 \times 3 \times 3$) are all powers of 3. Two of the largest powers are called googol (10 multiplied by itself 100 times, or 10^{100}) and googolplex (10 multiplied by itself a googol number of times, or 10^{googol}).

Powers are used in virtually all areas of business, science, and education from helping schoolchildren with their studies of mathematic and defining the magnitude of earthquakes, to analyzing the amounts of acid rain that fall on manufacturing areas.

Fundamental Mathematical Concepts and Terms

A compact notation has been developed to represent powers. The notation is described by a^n, where the first variable a is the base that is successively multiplied by itself and the second variable n is the exponent which indicates the number of times the base is to be multiplied by itself.

The powers of 10 are: 10, 100, 1000, etc., represented as 10^1, 10^2, 10^3, etc. Many powers of ten have been given distinct names. For example, in the United States, thousand denotes 10^3, million represents 10^6, and billion stands for 10^9. As one example, 10^6 equals $10 \times 10 \times 10 \times 10 \times 10 \times 10$. When these six numbers are multiplied together, the result is 1,000,000 (or written out as one million). The definition of powers can be broadened to include zero, negative, and rational exponents. Powers are also used within scientific notation, logarithms, and series.

A Brief History of Discovery and Development

Using powers of numbers originated with the ancient mathematicians, most likely with the Egyptians (whose civilization lasted from about 3300 B.C. to 30 B.C. and the Babylonians (living from the eighteenth century B.C. to the sixth century B.C.). These ancient peoples encountered powers when developing formulas to

describe geometric forms. The Egyptian Rhind pyrus, which dates from 1650 B.C., contains the concept of powers for numbers. The Pythagoreans (c. 450 B.C.) originated the use of x-squared for x^2 and x-cubed for x^3.

Diophantus of Alexandra (c. 200–284) used S for the square of an unknown, C for the cube, SS for the square square (fourth power), SC for the square cube (fifth power), and CC for the cube cube (sixth power). Ways to represent powers of unknowns began to spread throughout many countries in the fifteenth century. For instance, French physician Nicholas Chuquet (1445–1488) denoted successive powers of an unknown by placing numerical superscripts on the coefficients. He represented $4\times^5$ as 4^5.

The first mathematician to use letters for numbers as a way to perform mathematical calculations is generally said to be Francois Vite (1540–1603), an advisor to King Henri IV of France. Vite used vowels (A, E, I, O, U, and Y) for the unknowns and consonants (such as B, G, and D) for known quantities. The convention where letters near the beginning of the alphabet represent known quantities while letters near the end represent unknown quantities was introduced later by René Descartes (1596–1650). Descartes also introduced the notation of x, xx, xxx, etc.-where today mathematicians prefer x, x^2, x^3, etc.

Real-life Applications

AREAS OF POLYGONS AND VOLUMES OF SOLID FIGURES

The areas of polygons and the volumes of solid figures are expressed as powers of a particular length of the figure. For example, the area (A) of a square of side s is calculated as side s times side s, or $A = s^2$; that is, the second power of s. The volumes of solid geometrical figures are designated as the third power of a length. For example, the volume (V) of a cube with sides of length y is calculated as $V = y \times y \times y$, or $V = y^3$, where y^3 is the third power of y. In the case of a sphere with radius r, its equation for its volume is $V = (4/3)\pi r^3$, represented as 4^π times the third power of r.

EARTHQUAKES AND THE RICHTER SCALE

Everyday more than one thousand earthquakes occur around the world. Most of them are not noticed because they originate beneath the ocean, far underground, or are too small for humans to feel. The surface of the Earth consists of large pieces, or plates, which constantly grind against each other. When sufficient pressure builds up beneath two plates, it is released through cracks, or faults, between the plates. The result is an earthquake, or shaking of Earth's surface.

In order to calculate the magnitude of earthquakes, American seismologist Charles F. Richter (1900–1985) developed in 1935 a scale (now called the Richter scale) for measuring earthquake strength. The amplitude of the waves caused by the energy released in an earthquake increases by powers of 10 with respect to the magnitude numbers used by Richter. The released energy of an earthquake can be approximated by an equation that includes the energy magnitude of these waves and the distance from the measuring device, called a seismograph, to the earthquake's epicenter. Numbers for the Richter scale range from zero to infinity, although nine is generally the top limit ever reached. The Richter scale grows by powers of 10, where an increase of one point means that the strength of that earthquake is 10 times greater than the level before it.

For example, the famous San Francisco earthquake of 1906 (when later evaluated by the method developed by Richter) had a Richter reading of 7.8, which is 10 (10^1) times more intense than one with a reading of 6.8, 100 (10^2) times more intense than one with a reading of 5.8, and 1,000 (10^3) times more intense than one with a reading of 4.8.

COMPUTER SCIENCE AND BINARY LOGIC

In order to store digital information on modern computers, such as on the memory of hard-drives, computer hardware is made up of millions, or even billions, of tiny switches that can be either turned OFF or ON. The digits, 0 and 1, are used to stand for these two states of OFF and ON, respectively. Since these switches have exactly two different values, computer scientists work with a numbering system based on two digits. That numbering system is called the binary number system, which uses 2 as its base number. Each digit in a binary number represents a power of 2 ($2^0 = 1$, $2^1 = 2$, $2^2 = 4$, $2^3 = 8$, $2^4 = 16$, etc.).

Computers have been designed to use two voltage levels—usually 0 volts for logic-0 and either +3.3 volts or +5 volts for logic-1. With these two voltage levels, computer scientists can represent the two different values OFF and ON or, equivalently, values such as no and yes, false and true, low and high, and many other combinations. Since only two digits are used, any binary digit, or bit (the smallest unit of information inside a computer), can be transmitted and recorded electronically simply by the presence or absence of an electrical pulse or current. Even though it takes many more digits to represent

binary numbers versus decimal numbers (for example, the decimal number 255 is represented in binary as 1111 1111), the greater speeds possible with the use of binary logic more than compensates for that fact.

ACIDS, BASES, AND pH LEVEL

Acidic and basic are two classes of chemical compounds that possess opposite characteristics. Acids are characterized as tasting tart, being able to change pink litmus paper to red, and often reacting with some metals to produce hydrogen gas, while bases taste bitter, turn litmus paper to blue, and feel slippery to the touch. Mixing acids and bases can cancel out their opposite characteristics, producing a substance that is neither acidic nor basic, but neutral.

In order to measure all the different chemicals found on Earth, the pH scale was developed to show how acidic or basic a substance is. The pH scale ranges from 0 to 14, with a substance having a pH of 7 considered neutral, one with a pH less than 7 being acidic, and one with a pH greater than 7 considered basic. The method of pH uses powers for comparing chemicals. Each whole pH value below 7 is ten times more acidic than the next higher value. For example, a substance with a pH of 4 is ten times (10^1) more acidic than a substance with a pH of 5, 100 times (10^2, or 10×10) more acidic than a substance with a pH of 6, and 1,000 times (10^3, or $10 \times 10 \times 10$) more acidic than a substance with a pH of 7. The same rationale is valid for pH values above 7, each of which is ten times more alkaline (basic) than the next lower whole value. For example, a substance with a pH of 10 is ten times (10^1) more alkaline than a substance with a pH of 9 and 100 times (10^2, 10 times 10) more alkaline than a substance with a pH of 8.

Knowing the value of pH is very important to many industries around the world. For example, the food industry relies on pH when dealing with all kinds of foods. The pH of carbonated colas (which contain phosphoric acid) is about 2.5, the pH of milk is about 6.5 (almost neutral), the pH of water is 7.0 (neutral), and the pH of bananas, garlic, and broccoli are all within the basic range.

The amount of pH in the atmosphere is important when acid rain falls on the Earth. Acid rain is a form of air pollution in which airborne acids, which are produced by electric power plants and other sources, fall to Earth in local and distant regions. Acid rain dissolves and washes away nutrients needed by plants, attacks trees, and damages bodies of water by making waters more acidic that then can harm fish and other aquatic animals. Because the corrosive nature of acid rain causes widespread damage to the environment, environmental scientists study acid rain in great detail. With an accurate measure of the pH of substances, based on the powers of numbers, scientists are better able to study and analyze the causes of acid rain and the ways to reduce or eliminate it.

ASTRONOMY AND BRIGHTNESS OF STARS

In astronomy, magnitude is a term used to designate the brightness of a star. The Greek astronomer Hipparchus (190 B.C.–120 B.C.) devised this system around 150 B.C. when he placed the brightest stars into the first magnitude class, the next brightest stars into second magnitude class, and so on until he reached the dimmest magnitude stars which were placed within the sixth magnitude class. By the nineteenth century, astronomers had developed the technology to objectively measure a star's brightness with the use of powers. Instead of abandoning the long-used magnitude system, astronomers modified it for their own use. They established that a difference of 5 magnitudes corresponds to a factor of exactly 100 times in intensity. For example, first magnitude stars are about $2.512^1 = 2.512$ times brighter than second magnitude stars, $2.512^2 = 2.512 \times 2.512$ times brighter than third magnitude stars, and $2.512^3 = 2.512 \times 2.512 \times 2.512$ times brighter than fourth magnitude stars, etc. Some very bright objects can have magnitudes of zero or even negative numbers and very faint objects have magnitudes greater than +6.

Potential Applications

THE POWERS OF NANOTECHNOLOGY

Powers is such a widely used term within mathematics that it will always be part of future applications. One promising new technology that will use powers in its development and application is nanotechnology, which is the research and development involved in manipulating materials on a very small scale so that microscopic machinery can be built. These nanotechnology materials and devices generally range from 1 to 100 nanometers, where one nanometer is equal to one-billionth of a meter (0.000000001, or 10^{-9} meter). Because scientists believe that nanotechnology will eventually give humans the ability to mold individual atoms and molecules into microscopic-sized biological, electrical, and mechanical machines, it may replace many current production processes.

Where to Learn More

Books

Berlinghoff, William P., and Fernando Q. Gouva. *Math Through the Ages: A Gentle History for Teachers and Others*, Expanded Edition. The Mathematical Association of

Key Terms

Decimal number system: A base-10 number system that requires ten different digits to represent numbers.

Logarithm: The power to which a base number, usually 10, has to be raised to in order to produce a specific number.

Scientific notation: A shorthand way to write very large or very small numbers.

America, Washington, D.C. and Farmington, ME: Oxton House Publishers, 2004.

Bluman, Allan G. *Mathematics in Our World*. Boston, MA: McGraw-Hill Higher Education, 2005.

Katz, Victor J. *History of Mathematics*, abridged edition. Reading, MA: Addison-Wesley, 2003.

Suzuki, Jeff. *A History of Mathematics*. Upper Saddle River, NJ: Prentice Hall, 2002.

Web sites

Bagenal, Fran, University of Colorado. "Algebra: Powers/Numbers, Variables, and Rules." Atlas Project. <http://dosxx.colorado.edu/~atlas/math/math61.html> (March 14, 2005).

Eames, Charles, and Ray Eames. "Powers of 10." <http://powersof10.com/> Eames Office. (March 16, 2005).

University of St. Andrews, Scotland "François Viéte." School of Mathematics and Statistics. January 2000. <http://www-groups.dcs.st-and.ac.uk/~history/Mathematicians/Viete.html>(March 14, 2005).

University of St. Andrews, Scotland "René Descartes." School of Mathematics and Statistics. December 1997. <http://www-groups.dcs.st-and.ac.uk/~history/Mathematicians/Descartes.html> (March 14, 2005).

University of St. Andrews, Scotland "Diophantus of Alexandria." School of Mathematics and Statistics. February 1999. <http://www-groups.dcs.st-and.ac.uk/~history/Mathematicians/Diophantus.html> (March 14, 2005).

University of St. Andrews, Scotland "Nicolas Chuquet." School of Mathematics and Statistics. December 1996. <http://www-groups.dcs.st-and.ac.uk/~history/Mathematicians/Chuquet.html> (March 14, 2005).

Prime Numbers

Overview

A prime number is a number that is larger than 1 and which can only be divided evenly by itself and by the number 1. Just a few examples of prime numbers are 2, 3, 5, 7, 11, 13, 17, 19, 23 and 29.

A Brief History of Discovery and Development

Prime numbers have fascinated people for centuries. When they were not battling Trojans and helping to devise philosophy and logic, the ancient Greeks were also tinkering with prime numbers. It was thought that these numbers held mystical power. The ancient Greeks were also interested in what came to be known as perfect numbers. These are numbers that can be divided evenly by other numbers (the divisors), with the divisors adding up to the original number. One example is the number 6. Six can be divided by 1 (to give 6), by 2 (to give 3), and by 3 (to give 2). Adding up the divisors $(1 + 2 + 3)$ equals 6.

Centuries before the modern era, mathematicians studied prime numbers. In 300 B.C., Euclid of Alexandria wrote an essay entitled 'The Elements' that collected the knowledge of mathematics up to that time. In 'The Elements', Euclid was able to demonstrate that prime numbers did not just stop at a predetermined value, but that they go on forever. In other words, prime numbers are infinite. Euclid also showed that if $2^n - 1$ is a prime number, then the number $2^{n-1} \times (2^n - 1)$ yields a perfect number.

Test Euclid's discovery by setting $n = 3$: $2^3 - 1 = (2 \times 2 \times 2) - 1 = 7$ (which is a prime number) so, $2^{3-1} \times (2^3 - 1) = 2^2 \times 7 = (2 \times 2) \times 7 = 28$. Twenty-eight can be divided into an even number by 1, 2, 4, 7 and 14; finally, $1 + 2 + 4 + 7 + 14 = 28$, so 28 is "perfect."

About 100 years later, another Greek mathematician, Eratosthenes, came up with a way of determining prime numbers. Among his other accomplishments, Eratosthenes was the first person to accurately estimate the diameter of Earth while serving as the chief librarian of the great ancient library in Alexandria. His prime-calculating invention was called the Sieve of Eratosthenes. This mathematical sieve drains away non-prime numbers from prime numbers.

To illustrate, Table 1 shows an arrangement of the numbers 1–100:

Perform the following steps:

- Cross out 1 (it's not a prime number)
- Circle 2 (the smallest prime number), then cross out every multiple of two (4, 6, 8, etc; in other words, every second number)

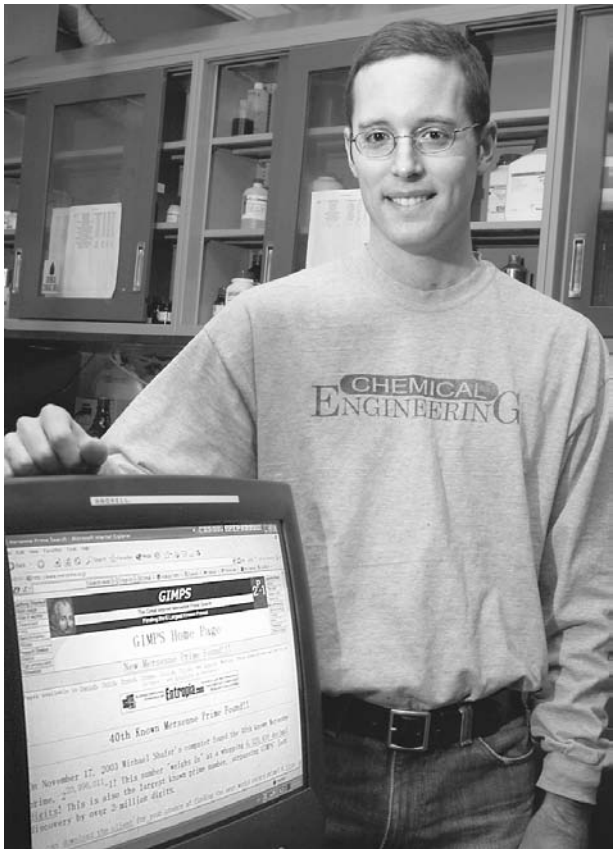
- Circle 3 (the next prime number) then cross out all the multiples of 3 (6, 9, 12, 15, etc.; some have already been crossed out)
- Circle the next number not circled or crossed out, which is 5, then cross out the multiples of 5 (10, 15, 20, 25, etc.; some have already been crossed out)
- Continue doing this until all the numbers have been circled or crossed out.

The circled numbers are the prime numbers.

Another prime number discovery made in the seventeenth century was made by Christian Goldbach, a historian and mathematician. He said that every even number could be expressed as the total of two prime numbers. As two examples, 6 can be expressed as 3 + 3, and 20 can be expressed as 17 + 3. His idea is known as the Goldbach conjecture. Even today, we are still not sure if his idea is true. But, scientists do know that the pattern is true for every even number between 2 and 400,000,000,000,000, and for some even numbers selected up to 10^{300} (10 followed by 300 zeros). In 2000–2002, a British firm offered a million dollars to anyone who could prove or disprove the Goldbach conjecture. No one did.

Since the 1700s, another a great challenge has been to determine the greatest prime number. Until the number-crunching power of big computers, this sort of activity did not get very far. However, with modern supercomputers the greatest prime number known now has over 4 million digits.

As recently as 2003, discoveries were announced regarding prime numbers. In that year, a team of physicists published a scientific paper in the prestigious journal *Nature* that provides evidence that the arrangement of prime numbers in amongst the other numbers is not just haphazard, but may have a pattern. Scientists and mathematicians are not sure what the significance of this might be. But, prime numbers may have a key role to play in the natural world.



Michigan State University graduate student Michael Shafer stands next to the computer he used to discover the world's highest prime number. The number is 6,320,430 digits long and would need 1,400 to 1,500 pages to write out. AP/WIDE WORLD PHOTOS. REPRODUCED BY PERMISSION.

1	2	3	4	5	6	7	8	9	10
11	12	13	14	15	16	17	18	19	20
21	22	23	24	25	26	27	28	29	30
31	32	33	34	35	36	37	38	39	40
41	42	43	44	45	46	47	48	49	50
51	52	53	54	55	56	57	58	59	60
61	62	63	64	65	66	67	68	69	70
71	72	73	74	75	76	77	78	79	80
81	82	83	84	85	86	87	88	89	90
91	92	93	94	95	96	97	98	99	100

Table 1.

Real-life Applications

BIOLOGICAL APPLICATIONS OF PRIME NUMBERS

Plant-eating insects called cicadas spend a lot of their life underground in one form, before emerging as adults. In some types (species) of cicada, this appearance occurs at the same time for all the adults in the region, every 13 or 17 years.

Thirteen and 17 are prime numbers. Coincidence? Scientists who have studied the species of cicadas do not think so. Rather, they think, the use of a prime number for the life cycle has been a response to the pressure put on cicada population by other creatures who utilize them as food. In other words, the cicadas are the prey and the

creatures lying in wait when they emerge to the surface are the predators.

Researchers have used mathematical ways to model the so-called predator-prey relationship. Modeling allows them to do experiments in their lab, on the computer, without having to actually go to nature and observe what is happening (which could be very hard to do).

In the mathematical model, the cicadas and their predators had life cycles that were randomly chosen to be different lengths. When both predator and prey were present in high numbers at the same time, it was bad news for the cicadas, as there were lots of hungry predators waiting for the cicadas as they came out of the ground. But, if the emergence of the cicadas occurred when there were not many predators, they had a much better chance of living long enough to mate.

In the computer studies, the researchers found that the best times for the cicadas to emerge from the ground was in life cycles that had prime numbers (e.g., 13 and 17 years). The researchers assert that a life cycle that is 13 or 17 years long increases the cicadas chances of avoiding population depletion. Consider what could happen if their life cycle was 12 years long. If cicada emerged every 12 years, any predator that had a life cycle of numbers that can divide into 12 (such as 2, 3, 4, or 6 years) could be around at the same time the cicadas emerged from the ground. There would be more chance of a hungry predator would be waiting. But, if a life cycle is 13 or 17 years long, a predator's life cycle also has to be 13 or 17 years long. The odds of that are much less.

Where to Learn More

Books

Cobb, C. *Cryptography for Dummies*. New York: For Dummies, January 2004.

Schneier, B. *Secrets and Lies: Digital Security in a Networked World*. New York: Wiley, August 2000.

Periodicals

Gole, E., O. Schulz, and M. Markus. "A Biological Generator of prime Numbers," *Nonlinear Phenomena in Complex Systems* (2000) 3: 208–213.

Web sites

Alfeld P. "Eratosthenes of Cyrene." <http://www.math.utah.edu/~alfeld/Eratosthenes.html> (September 7, 2004).

——— "Notes and Literature on Prime Numbers." <http://www.math.utah.edu/~alfeld/math/prime.html> (September 6, 2004).

Peterson I. "Prime-Time Cicadas." <http://www.sciencenews.org/articles/20030621/mathtrek.asp> (September 8, 2004).

Probability

Overview

Probability is the likelihood that a particular event will occur. Probability is used to estimate the chances of many different types of events happening. Insurance companies use probability to estimate how likely a particular driver is to cause an accident during the next year. Engineers use probability to predict how often critical pieces of equipment, such as jet engines on passenger planes, will fail. Gamblers in casinos routinely make wagers based on their understanding of the laws of probability, while investors make even riskier gambles on the rise and fall of the stock market or the price of a bushel of corn. Although probability is one of the most commonly used forms of mathematics in everyday life, many misconceptions exist about its formulation, meaning, and impact.

Fundamental Mathematical Concepts and Terms

Probability calculations are generally straightforward, though as the number of possible outcomes grows, the math required can become somewhat involved. Consider a simple example involving a single die (the singular form of dice), in which we wish to determine the probability of rolling a 4. The calculation for probability includes several elements. Outcomes are all the possible results we could achieve; since the die has 6 sides (1, 2, 3, 4, 5, and 6), and any of the six could land on top, the total number of possible outcomes in this experiment is 6.

Next, we must determine the total number of ways in which the event of interest could possibly occur; in this case, a roll of 4 can occur only one way. By dividing this value (the number of ways our desired outcome can possibly occur) by the total number of possible outcomes, we can determine the probability of the 4 being rolled, creating this equation: Probability = Desired outcome / Total Outcomes, or in numerical terms, $P = 1/6 = 1/6$. Thus we conclude that the probability of rolling a 4 on a single toss of the die is 1/6, or 1 in 6. We could perform the same calculation for each of the other values on the die, demonstrating that for each side of the die, the probability is also 1 in 6.

Interpreting this value is relatively straightforward: a probability of 1 in 6 tells us that if we roll the dice a large number of times, we will, on average, roll a single 4 for each six tosses of the die. If we wish to find out about how many 4s we will roll in 600 rolls of the die, we multiply the probability by the number of rolls, which are often called experiments or trials; in this case, we use the following

equation: $1/6 \times 600 = 100$. This result tells us that over the course of 600 rolls, about 100 will be 4's.

This same procedure can be scaled up to evaluate events with thousands or millions of possible outcomes. If the names of each person living in the U.S. were written on slips of paper and one slip was randomly drawn, what chance would John Smith of Cloverleaf, Iowa, have of being drawn? In this example, only one John Smith exists in Cloverleaf, providing only one possible way to reach the desired outcome. The total number of people in the U.S. in 2005 was approximately 300,000,000, which is the total number of possible outcomes. John Smith's chance of having his name drawn is 1 in 300,000,000. If the drawing were held using the earth's entire population of 6,400,000,000, John's chance of being drawn would drop by a factor of 20.

A Brief History of Discovery and Development

While the very first game of chance cannot be specifically identified, historians are certain that these probability contests have been enjoyed for millennia. Ancient civilizations left behind small dice-shaped pieces of bone called astragalia, which apparently facilitated the earliest contests similar to modern dice games. Throughout the early history of man, gambling remained popular, with little apparent attention paid to the laws of nature and mathematics, which made the toss of the colored stones or polished bone fragments so maddeningly unpredictable.

During the sixteenth century, Gerolamo Cardano (1501–1576), a scholar of medicine, astrology, and philosophy made the first known attempt to explain the function of chance in gambling and other endeavors. Cardano was the first to deduce that an event's probability of occurring is determined by dividing the number of ways the event could occur by the total number of possible outcomes. Cardano explained that a roll of a single die has six possible outcomes, while a pair of dice can land thirty-six different ways; he also wrote about the statistical logic of a primitive ancestor of the modern game of poker. Unfortunately, Cardano's science was somewhat limited by his intense belief in astrology, which he used to predict future events of human lives. Perhaps his most successful prediction was naming the date of his own death far in advance; when the predicted date of his death arrived, Cardano insured his own correctness . . . by committing suicide.

A century later, mathematician Blaise Pascal (1623–1662) was asked why the odds of throwing a single six in four throws of one die do not equal the odds of throwing four sixes in twenty-four throws of two dice. Pascal accepted this challenge, then went on to devise the theory of probability as it is currently understood, in many cases applying these principles to the popular pastime of gambling. At age nineteen, Pascal also constructed the first mechanical adding machine.

While various other mathematicians added to the body of knowledge regarding chance and gambling in the decades that followed, the next major advance occurred in 1928, when John von Neumann put forward the basic concepts of game theory in a paper analyzing the probabilities associated with various poker hands. While game theory has found application in fields such as economics, its application to games of chance also continues, particularly given the advent of powerful, inexpensive computers.

Real-life Applications

SECURITY

The ability to conceal data from outsiders has been valued by military commanders for centuries; historians have uncovered evidence of military codes dating back more than 4,000 years. The Allied victory in World War II was hastened significantly when the Allies broke a presumably unbreakable code used by the Japanese, thus becoming privy to numerous confidential communications. Today, numerous applications for encoding and decoding data exist, most of them based on the fundamental principles of probability.

One critical use for this technology is data encryption, a technique for encoding data so that it is unreadable without a specific number, or key, which allows an authorized user to decrypt and read the message. Data encryption has become a critical technique as electronic transfers of sensitive financial data have become more routine. Many commercial websites now transfer buyers to a secure site, at which data such as credit card numbers is encrypted before it is transmitted from a user's computer.

Encryption works because of the laws of probability. An encrypted message can be read by any person with the proper numerical key, meaning that for a message to remain secure, the key must be virtually unguessable. Ever faster computers have made it possible for simple encryption schemes to be broken using a brute-force approach, in which the computer simply tries key after key until the proper one is located. Preventing this type of attack requires a large enough number of possible keys that the likelihood of guessing the proper key by chance becomes so small that it is not worth attempting. An

encryption key's resistance to brute force attacks is measured as strength, with a more secure key being described as stronger encryption.

As of 2005, one of the most widely used encryption schemes is found in Microsoft Internet Explorer, where it encrypts data sent from computer users to the Internet. This encryption scheme uses a 128-bit encryption key, meaning that in order to read the encoded data, an interloper would have to correctly guess a 128-bit number. Since this length of key would theoretically take a modern supercomputer several hundred years to crack, 128-bit encryption is considered adequate for routine applications such as online shopping. In cases where additional security is desired, such as military applications, longer keys significantly increase the number of possible keys, producing a commensurate reduction in the odds of randomly guessing the key.

A related use for encryption techniques has recently appeared in the rapidly growing field of forensic computing. In the course of criminal investigations, law enforcement personnel frequently need to locate computer files related to a crime, a process much like finding the proverbial needle in a haystack. A typical computer hard drive contains hundreds of thousands of files, most of which arrive as part of the operating system or are installed with user applications; a basic installation of Microsoft Windows XP places between 10,000 and 30,000 separate files on a computer hard drive.

Unlike a computer user who knows where most of his important files are saved, a police investigator searching a computer for files with evidentiary value has no idea what the needed files are called, or in which directories they reside. Since it is impractical to manually open and read every file on the computer, encryption methods now allow investigators to automatically eliminate more than 90% of the files on a computer, permitting the investigator to focus on the remaining files.

This file-sorting system is based on the principle of encryption, in which any file can be processed to produce a unique identifying code. By creating these unique codes, or file signatures, of all the files installed by most operating systems and commercial applications, investigators have created a massive reference library for law enforcement purposes. Investigators can use this library to scan a suspect's hard drive, automatically eliminating any files which match the signature keys of known files while leaving the files which might have evidentiary value behind. The system works only because the number of potential file signatures is enormous; in the case of the MD5 algorithm, the total possible number of unique file signatures is 10^{38}, or a one with 38 zeros after it, making the odds of two files having the same file signature almost an impossibility. By reducing the number of files to be examined, this library enables investigators to more rapidly and more efficiently search hard drives, gathering evidence they might otherwise overlook.

GAMBLING AND PROBABILITY MYTHS

While the ancestors of today's dice games predate recorded history, the modern game of craps is far more recent, and is attributed to twelfth century Crusaders besieging a castle in Arabia. Most of today's other casino games also can be dated back to the Middle Ages, however one type of wagering can rightfully trace its lineage back more than twenty centuries. The longest-running wagering event practiced today is the ancient sport of horse racing.

Numerous archaeological finds support horse racing's claim as the most ancient form of gambling. A Hittite document dating to about 1500 B.C. describes in detail the process of breeding and training horses for the purpose of racing, while the Iliad provides a complete account of a chariot race. The Olympic games in 624 B.C. included specific rules for horse racing in contests of various distances, and the Romans soon added the concept of handicapping, or betting against the house. While the popularity of horse racing has risen and fallen over the centuries, today's racing, while faster and more refined, is virtually unchanged from the ancient contests held in Europe. While the advent of modern statistical analysis and computer equipment has provided the tools to analyze the mountains of statistical data available on past races, the ability to correctly predict the outcome of a horse remains an elusive goal.

While the interpretation of probability projections is fairly straightforward when applied to events which occur many times, the laws of probability become far less intuitive over short periods of time. One common probability myth, often cited by gamblers, is that numbers, horses, or players can become due, meaning that since they have not won in many plays of the game, they are now more likely to occur. This faulty line of reasoning is based on the understanding that over many thousands of plays, each number will appear a set number of times, hence the gambler assumes that the longer a value goes without appearing, the more likely it is to appear soon. Unfortunately, this belief is unfounded. In the case of completely unrelated events, such as the spin of a roulette wheel, the odds of the next spin are unchanged by the result of any previous spins. If the number 14 has not been spun on a particular wheel for six weeks, the odds of it appearing on the next spin are still exactly the same as they were before. The laws of probability do not provide for events to occur simply because they have not occurred previously.

The probability of this coin landing heads or tails is easy to predict. ROYALTY-FREE/CORBIS.

A second probability myth, ironically, is the exact opposite perspective of the previous view. This perspective says that particular numbers can become "hot," or more likely to be spun. In adopting this philosophy, an observant gambler might notice that the number 27 had been spun on the wheel several times over the course of a short wagering session. The gambler, acting on the theory that numbers can become hot, now concludes that the number's frequent appearance in past spins makes it more likely to appear in a future spin, and he will wager heavily on this particular number. Once again, the laws of probability and chance dictate that, assuming the roulette wheel is functioning correctly, the chance of a future spin cannot be predicted by how often a particular number has appeared in recent spins. Regardless of how hot a number appears, it is no more likely to appear on the next spin of the wheel than any other value. Ironically, the theory of hot numbers, which says that the same number will come up many times together, is the exact opposite of the theory of coming due, which says that a number will appear when it has not been spun for some time. While gamblers subscribe to both philosophies (and back up their philosophies with their wallets), both theories cannot simultaneously be right; in truth, probability theory says that neither theory is correct, and that past events do not impact future spins of the wheel.

PROBABILITY IN SPORTS AND ENTERTAINMENT

Many sports rely on probability to predict future events. Baseball is among the most statistically-oriented sports, with numbers available for almost every aspect of the game. A player's batting average is a measure of the percentage of times he hits safely, expressed as a 3-place decimal value such as .333. While this value allows an assessment of a player's past performance, it is also useful in predicting his future effectiveness. For instance, a player batting .200, which can also be expressed 2:10, 1:5, or 20% can be predicted to hit safely 20% of his times at bat, or 1 time in 5 attempts. For this batter, the odds against him hitting safely on any given trip to the plate will be 4:1. Baseball batting averages are calculated using an involved set of rules, meaning that a player batting .200 will generally make it to first base safely more than 20% of the time; for this reason, some managers prefer to use a player's on-base average, which includes walks and errors in the player's success ratio.

Bowling is a popular sport in which players actually receive two chances to succeed, in the form of two shots (if necessary) to knock down all ten pins. Statisticians have used mountains of data from previous bowling competitions to calculate the odds of a professional bowler making a variety of shots. For example, when a professional bowler steps up to roll his first ball in a frame, the objective is to knock down all ten pins, scoring a strike. For the second shot, the odds of clearing the lane depend on which pins remain standing; three pins standing close together have much higher odds of falling than two widely separated pins. For most bowlers, a split is one of the hardest shots in the game, requiring the player to slice the ball to the outside of one pin, knocking it across the lane to hit the other one. Even for a professional, splits are long-shots. According to the Professional Bowlers Association, the 7-10 split, in which two pins remain at opposite sides of the lane, has been attempted 400 times in televised matches. In all these attempts, the professionals have managed to convert only three, putting the odds of a professional making this shot at 3:400, or about 1 time in 133 attempts.

How often do miracles occur? The term miracle has several different meanings; in theological language, it refers to an act of God that defies the laws of nature,

though its most common use today refers to any seemingly impossible event that actually occurs. When the Boston Red Sox finally broke the decades-long curse of the Bambino and won the World Series in 2004, fans proclaimed the victory a miracle. When a jet airliner crashes and one or two passengers walk away without injury, many label their survival a miracle. And in a handful of cases where a single individual has won a state lottery, not once but twice, writers routinely throw out the term to describe this odds-defying run of luck. Ironically, the term is applied almost exclusively to positive events like those described, ignoring equally improbable turns of probability which lead to unexpected death or injury.

While no statistical definition of miracle exists, an estimate can be made based on common language. To most people, the expression "one in a million" describes something quite rare, though still achievable. People frequently use this expression to describe a job they truly love or a dear family member or friend, suggesting that this level of probability does not rise to the level of miracle status. For this discussion, we must conclude that a miracle is much rarer than 1 in a million; for simplicity's sake, we will assume that a probability of 1 in one billion qualifies an event as a miracle. In other words, a miraculous event is one which occurs only once in every billion opportunities. To get some sense of this level of probability, one billion seconds would take more than 30 years to elapse.

In determining how often these miraculous events occur, it becomes important to recognize that while odds of 1 in one billion are almost unimaginably low, these odds apply not just to a single person, but to many millions of individuals. For example, assume that any single person in the United States has a miraculous, or one in a billion, chance of being struck by lightning in a given day. With odds like these, any single person can safely go on with his life without worrying about storm clouds. But when these odds are applied to the entire 300 million people in the U.S., the equation changes dramatically since each of the 300 million provides another opportunity for the miraculous event to occur. Now, across the entire population, the odds of a lightning strike in a day become 300 million in one billion, or roughly 1 in 3.33. At this probability level, some individual in the U.S. would be struck by lightning every three days, making the miraculous seem almost routine, since many of these strikes would undoubtedly be covered on national news. Fortunately, lightning strikes appear to be infrequent enough to reach even the so-called miraculous level proposed here. But given the large number of citizens in the U.S., it seems statistically likely that one in a million events actually occur on the North American continent several times each day.

PROBABILITY IN BUSINESS AND INDUSTRY

Some business endeavors require a calculation of probabilities, even though little data on which to base the calculation is available. Complex pieces of machinery like the NASA space shuttle are notoriously hard to estimate reliability projections for, due largely to the massive number of components involved. Some components are simple; for example, a tire on the space shuttle is one of the more dependable components. Other components contain thousands of parts; the shuttle's main engines are among the most complex propulsion systems ever designed. In order to calculate the odds of an accident occurring in a single shuttle flight, the chances of failure for each individual component must be calculated, then combined with those of the other components to produce a composite estimate of the ship's chances of returning safely.

As the number of components rises, the process becomes increasingly difficult; because of the shuttle's complexity this process becomes virtually impossible to carry out accurately for such a machine, sometimes forcing engineers to make an educated guess. Unfortunately, these guesses are sometimes given more credibility than they deserve. Prior to the shuttle *Challenger's* loss on the twenty-fifth shuttle mission, engineers had assessed the shuttle's chance of a catastrophic failure at 1 in 100,000, meaning the ship could have flown every day for 300 years while suffering only one major failure during that time. Unfortunately, these overly optimistic assessments appeared to ignore previous experience with unmanned solid rockets, which suggested an accident rate closer to 1 in 25 or 1 in 50 for the boosters alone. To date, actual experience with the shuttle system has led to 2 shuttle accidents in 113 missions, suggesting that the probability of loss is far closer to the 1 in 50 value than the 1 in 100,000 estimate.

Most consumer products sold today include a warranty period, during which the manufacturer agrees to either repair or replace the product if problems occur. For most products, users expect the item to last far beyond the warranty period; new automobiles typically include a three to five year warranty, even though most buyers expect to drive a new car for twice that long. In some cases, manufacturers attempt to estimate the likely service life of a product by providing a measure called mean time before failure, or MTBF. For example, a computer monitor might be sold with an advertised MTBF of 50,000 hours, which equates to 10 hours of use, 5 days per week, for more than nineteen years. For most customers, nineteen years is longer than they typically keep a monitor, so they will feel comfortable with this purchase. However,

MTBF is not the same as a warranty or a minimum lifetime; rather, MTBF provides the mean, or average lifetime of this product model before failure. In other words, half of the products will last longer than MTBF (50,000 hours in this case), but the other half will fall below the average, failing at some point less than the advertised lifetime.

If MTBF does not give a minimum lifetime, how should it be interpreted when trying to assess a product's potential service life? First, if the MTBF has been correctly calculated, the buyer can expect that the item will provide the rated service life or more half the time, so if he buys twenty of the monitors, he can expect at least ten to last 50,000 hours or longer, in some cases perhaps much longer.

The other monitors in the group can be expected to last for varying periods of time, with most of them lasting close to the average lifespan of 50,000 hours and a few failing as the time-span grows further from the mean. In a few cases, monitors might actually quit working within the original warranty period, meaning they would be replaced by the manufacturer. Unfortunately, MTBF calculations for complex electronic equipment can be impractical or impossible to calculate mathematically, meaning that in some cases the MTBF is based largely on engineer intuition and experience with similar parts, rather than actual experimentation.

One of the most exciting moments in a teenager's life is when she finally receives her driver's license. But soon after this triumph may come a rude surprise: car insurance for young drivers is often several times as expensive as for older adults. Why do insurers charge teen drivers more?

Insurance companies are among the largest users of statistical and probability data. Specialists called actuaries spend their days determining exactly how likely events are to occur, allowing the insurer to charge correctly for its policies. Actuarial tables provide summaries of this data; for example, an actuary could use one of these tables to determine that a 45-year-old man in good health is likely to live to be 82 years old, and that his odds of dying next year are 1 in 14,400. Using these probabilities, the insurer can then determine how much to charge the man for a life insurance policy which pays $100,000 to his family in the event of his death.

These probability tables allow insurers to provide discounts to specific customers, such as those who don't smoke, since they have a higher probability of living longer. Automobile insurers also use actuarial data to predict which drivers are more likely to be involved in an accident, in which case the insurer will be obligated to pay for repairs. Using this information, insurers then give lower rates to drivers who have lower odds of having an accident and higher rates to those with higher odds. Based on past experience with millions of drivers, insurance companies know that the odds of a teenage driver, especially a male, having an accident are much higher than for a 30- or 40-year-old. Since the company is more likely to pay a claim for these young drivers, it is forced to charge higher premiums in order to cover the expected losses. As long as young drivers continue to have more accidents in general, even safe teenage drivers will continue to pay higher premiums for auto insurance. In a few cases, actuarial data has shown that certain groups, such as Honor Roll students, are less likely to have accidents, and some insurers now offer discounts to students with strong academic performance.

OTHER USES OF PROBABILITY

In 2001 Russian engineers fired braking rockets to bring the aging Mir space station back to earth. The reentry was carefully orchestrated to insure that most of the station would burn up in the earth's atmosphere, and any surviving pieces would land harmlessly in the Indian Ocean. Recognizing the incredibly long odds of losing, restaurant chain Taco Bell made an astonishing offer. The company floated a 40-foot square target featuring the words "Free Taco Here" in the Indian Ocean off the coast of Australia. The company then widely advertised that if the remains of the Mir station hit the target, Taco Bell would give one free taco to every person living in the United States. Mir eventually landed thousands of miles from the target, and the company avoided having to serve 300 million free tacos. However, executives at Taco Bell apparently recognized that even the unlikeliest of events occasionally occurs; the company took out an insurance policy in advance just in case the falling station defied the exceptionally long odds and hit the target.

Sometimes the seemingly impossible can be accomplished due to an audience's lack of statistical savvy. Consider this simple magic trick. A magician, claiming to have psychic powers, stands before a crowd and announces that he has noticed an odd coincidence: although there are 365 possible birthdays in a year, he has psychically observed that two of the individuals in this particular audience happen to share the exact same birthday. He then asks a series of questions to help locate the unlikely pair, and after confirming this fact, moves on with his act. Was it psychic power, or simple probability?

To most casual observers, the large number of possible birthdays seems to make the prediction a long shot at best. But considered in terms of probability theory, it begins to look far less magical. Assume that the crowd consists of 12 people. The magician has a 0.5073 chance of

Key Terms

Actuary: A mathematical expert who evaluates the statistical likelihood of various insurable events for underwriting purposes.

Encryption: Using a mathematical algorithm to code a message or make it unintelligible.

Gambling: A popular form of entertainment in which players select one of several possible outcomes and wager money on that outcome.

Probability: The likelihood that a particular event will occur within a specified period of time. A branch of mathematics used to predict future events.

being correct, one better than in two. With a bit of showmanship, most psychic performers are able to easily dismiss the predictions they miss using a variety of explanations. But in close to half of this performer's appearances, he will shock the crowd by appearing to do the impossible, when in fact he has simply made a smart bet based on the simple laws of probability.

In a few instances, probabilities are used to attract attention or create fear. Newspaper and magazine headlines during the mid-1990s warned air travelers to avoid planes with fewer than thirty seats, based on statistics which seemed to indicate that these smaller planes were several times more likely to crash than larger jets. But this probability was based on a classification system which grouped small commercial planes in the same category as helicopters and some other types of planes, unrealistically inflating the numbers for the category and making the commuter planes seem less safe. Eliminating the other types of equipment from the equations produced probability figures demonstrating that the smaller commercial planes are approximately as likely to crash as their larger cousins.

Potential Applications

As computational power continues to double every two years, the ability to apply probability theory in new ways will lead to further applications for this powerful tool. In some cases, these applications may involve major improvements in current applications, such as forecasting weather patterns or predicting when and explaining why freeways suddenly become congested. The ability of faster computers to crack increasingly complex codes will lead to an escalating battle between code-writers and code-breakers.

In other cases, advances in probability theory may well result in unforeseen applications. Based on mathematical advances made by eighteenth century mathematician Thomas Bayes, scientists are just beginning to develop software which is comfortable dealing with concepts such as "probably" and "more likely" rather than the simple yes or no typically required in computer programming. Google and other search engines already use rudimentary forms of Bayesian reasoning to answer search queries. Potential future applications include cameras which would visually examine a patient and warn a physician of symptoms making the person more likely to suffer a stroke.

Where to Learn More

Books

Epstein, Richard A. *The Theory of Gambling and Statistical Logic.* New York: Academic Press, 1977.

Glassner, Barry. *The Culture of Fear; Why Americans are Afraid of the Wrong Things.* New York: Basic Books, 1999.

Orkin, Mike. *What Are the Odds? Chance in Everyday Life.* New York: W.H. Freeman and Co., 2000.

Web sites

Feynman, Richard P. "Personal Observations on the Reliability of the Shuttle." <http://www.virtualschool.edu/mon/SocialConstruction/FeynmanChallengerRpt.html> (March 27, 2005).

Singh, Simon. "Crypto Q & A." <http://www.simonsingh.com/Crypto_Q&A.html> (March 28, 2005).

Proportion

Overview

Proportion is an equation used to compare the magnitudes of quantities. It can be defined as an equation that presents equality between two ratios. In other words, if the ratio between two characteristics of an object is equal to the ratio between the same two characteristics of another object, the two objects are considered to be in proportion. These characteristics could be anything that can be measured (such as size, quantity, dimension, etc.). For example, consider two rectangles, the first having length and width equal to 8 in. (20 cm) and 4 in. (10 cm) respectively, and the other having a length of 6 in. (15 cm) and width of 3 in. (7.5 cm). These two rectangles are in proportion as the ratio of the length and width of each rectangle is equal.

Although proportion is a concept mainly used in design, it is widely applied to other aspects of daily life as well. One of the most common examples is grocery shopping, where proportion is frequently used to compare prices of items with different sizes. In addition, proportion finds uses in numerous other fields, including architecture, art, maps, astronomy, business, imaging, technology, and even cooking.

Fundamental Mathematical Concepts and Terms

As stated earlier, proportion is indicated by the equality between two ratios. Mathematically, it can be expressed in two ways—a/b = c/d or a:b = c:d. The outer terms of the equation are known as extremes, while the inner (or middle) terms are known as means. For example, in the above equation "a" and "d" are extremes, whereas "b" and "c" are means.

SOLVING RATIOS WITH CROSS PRODUCTS

One way to test equality is by simply calculating the values of the ratios. However, a more commonly used method involves the use of cross products. Cross products can be calculated by multiplying the outer terms (or extremes) and then the inner terms (means). If both values are equal, the ratios are in proportion.

Consider the ratios 2/5 and 3/7.5. In this case, the cross product of the extremes is $2 \times 7.5 = 15$, while the cross product of means is $5 \times 3 = 15$. Hence, the ratios are in proportion. Note that simple division here would have been far more complex and time consuming, as compared to calculating cross products.

This is one of the reasons for the popularity of the cross product method.

The cross product method also has another significant benefit. Real life applications use the concept of proportion mainly to compare two things or objects. In many cases, there may be a missing term in the proportion. For example, a grocery store owner charges $1.50 for 1 lb. (0.4 kg) of beef roast. He wants to set the price of a 3 lb. (1.2 kg) roast, such that it is in proportion with the price of the 1 lb. (0.4 kg) roast. This can be easily done by writing the proportion equation, and then using cross product to determine the price.

The equation can be written as—$1.50/1 = x/3$, where "x" is the price of the 3 lb. roast. By calculating the cross products of the means and extremes, the value of "x" comes out to be 4.50. In other words, the 3 lb. (1.2 kg) roast should be priced at $4.50 for it to be in proportion. Simply put, you can calculate a missing term from a ratio if this ratio is in proportion to another known ratio. This underlying concept of proportion is extremely useful in real-life applications.

DIRECT PROPORTION

If change in one component causes a change of equal magnitude (size, percentage) in another component, the two components are said to be in direct proportion. Another way of expressing this is by stating that the first component is directly proportional to the second component. In a nutshell, direct proportion is a concept that pertains to the change in the values of two (or more) components that are already in proportion.

For example, imagine the price of a candy bar is $0.50. The number of candy bars is always proportional to the total price of the bars—the ratio of the number of candy bars to the total price always remains same. One bar costs $0.50, two bars cost $1.00, four cost $2.00, eight bars cost $4.00, and so on. Put simply, a change in the number of bars causes a change in the total price. Moreover, the magnitude of the change is also the same. In other words, the change in the number of bars as well as the price can be represented by a common factor. The number of bars keeps doubling (or $1 \times 2 = 2$, $2 \times 2 = 4$, $4 \times 2 = 8$). Similarly, the price also doubles ($0.50 × 2 = $1.00, $1.00 × 2 = $2.00, $2.00 × 2 = $4.00). Hence, the number of candy bars is directly proportional to the total price of the bars. Also the change is represented by a common factor (two in this case).

Mathematically, direct proportion is indicated as $a \propto b$ (a is directly proportional to b). The main advantage of direct proportions is that they can be expressed in the form of an equation. For example, the relationship between the total number of bars and the total price, in the above case, can be shown as:

Total number of candy bars = k × Total Price, where k is the common factor.

The common factor is known as the proportionality constant. This equation may be used to easily calculate the total price if the number of candy bars is known, and vice versa. All direct proportion relationships can be expressed by such equations. Consequently, they are used extensively in various real-life activities and applications.

INVERSE PROPORTION

Like direct proportion, inverse proportion also pertains to the change in two (or more) components. However, in the case of inverse proportion, an incremental change in one component causes a decrement in the other component. In other words, if the magnitude of one component increases, the value of the other component decreases, and vice versa.

Consider, for example, a car traveling from one place to another. If the car has a constant speed (and assuming it does not stop anywhere), the more it travels, the less the remaining distance to the target destination. Hence, in this case, as the total travel time increases, the distance to the destination decreases—travel time is inversely proportional to distance remaining.

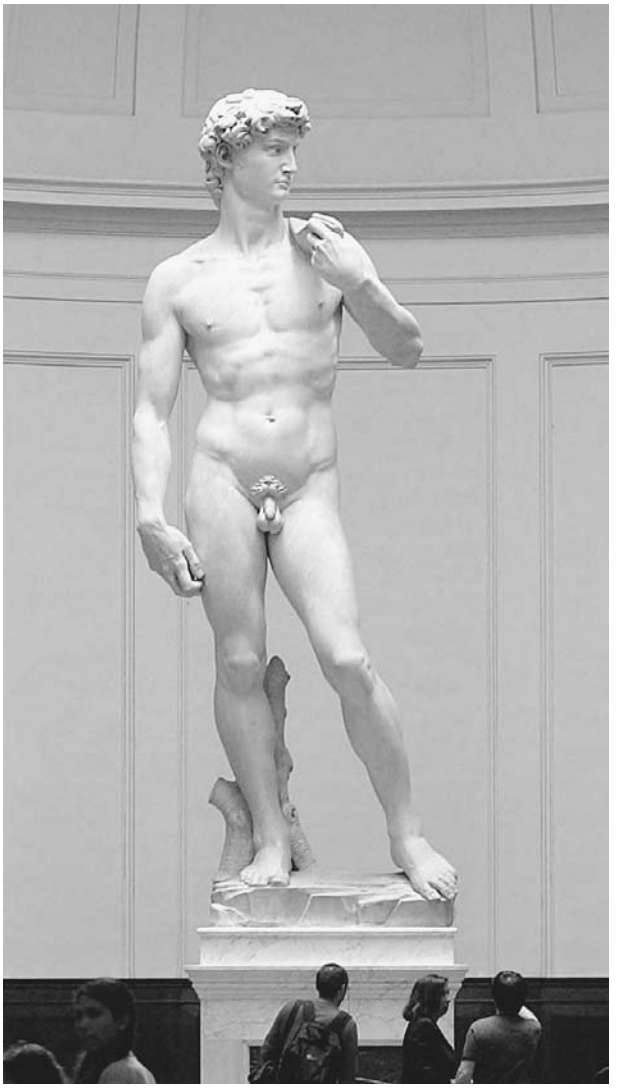
Similar to direct proportion, the change can be represented by a factor. However, the factors that represent change for both components are multiplicative inverses of each other. In simple terms, if the value of one component changes by a factor of three, the change in the value of the other component will be 1/3. Consequently, inverse proportion is also known as reciprocal proportion, and is mathematically indicated as $a \propto 1/b$ (or travel time $\propto$ 1/distance remaining, for the above example).

An inverse proportion relationship can also be expressed in the form of an equation. For instance, the two components (travel time and distance remaining) in the above example can be shown as:

Travel time × k/distance remaining, where k is the proportionality constant.

A Brief History of Discovery and Development

Throughout history, proportion has been used extensively in numerous areas. The Greek mathematician Pythagoras (580 B.C.–500 B.C.) who is most well known for the Pythagorean theorem, developed the Theory of



Michelangelo's marble statue of *David* in Florence, Italy. Measurements of the statue debunked long-held notions that the 13.5-ft (4.1-m) high statue was out of proportion to the human form. AP/WIDE WORLD PHOTOS. REPRODUCED BY PERMISSION.

Proportion to relate music with mathematics. He established musical scales that were based on the concept of proportion.

Subsequently, evidence of proportion can be seen in many works of art and architecture, especially in ancient Greece and Rome. Some of the most popular paintings by renowned artists such as Michelangelo (1475–1564), Raphael (1483–1520), and Leonardo da Vinci (1452–1519) were based on proportion. The concept of proportion is vital to art and architecture as it describes the size, location, or amount of one element to another within the entire work (e.g., *Vitruvian Man* by Leonardo da Vinci). The proportion of various parts of the body in this painting is very similar to the proportion seen in an average human body.

Similarly, much like modern architecture, ancient structures and buildings also incorporated proportion. The ancient Egyptians used it in the construction of the pyramids. The Parthenon in Athens, Greece, is another structure where proportion, along with ratio and scale is used extensively to create a "harmony" among various elements.

Interestingly, Isaac Newton's (1643–1727) second law of motion states that the acceleration of an object in motion is directly proportional to the force applied on it—a classic equation indicating direct proportion between two properties, acceleration and force.

Historians and mathematicians also believe that the great musicians Mozart (1756–1791) and Beethoven (1770–1827) used proportion to compose music. Proportional scaling allows the composition of harmonic, pleasant-sounding, music—a concept initially put forward by Pythagoras.

Subsequently, by the nineteenth century proportion was applied to numerous applications including those in business and sciences.

Real-life Applications

ARCHITECTURE

Architecture uses mathematical concepts such as proportions and ratio extensively. Since ancient times, architects and designers have been building various parts of a structure in proportion to attain visual appeal, unity, stability, and order. These principles hold true even today. Proportion is employed in a number of ways in architecture. Most popular buildings and structures—ancient as well as modern, are based on what is commonly known as the divine proportion or golden proportion.

The divine proportion consists of two or more ratios that are equal to phi (or 1.618). In other words, if the ratio (also known as divine ratios) of various parts of a building (or a structure) is equal to the number 1.618, then the proportion of these various parts is known as the divine proportion. Throughout the world, monuments, famous buildings, and other structures have been created using the divine proportion. This includes the pyramids of Giza, the Parthenon in Greece, the Colosseum in Rome, numerous cathedrals including St. Peter's Cathedral in the Vatican, the Taj Mahal in India, the Pentagon in the United States of America, and many more.

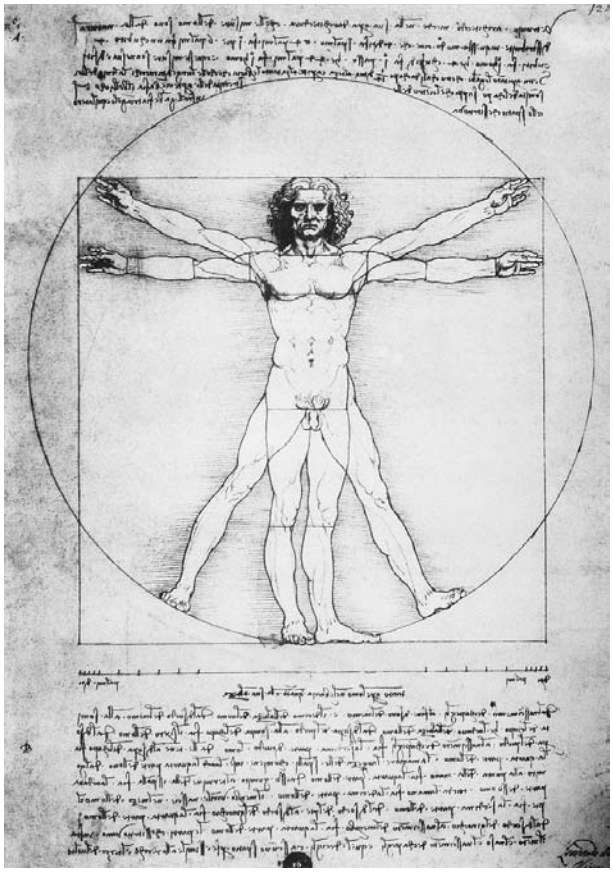
As stated earlier, proportions are used on various elements (or parts) of the entire structure. For example,

the front elevation of the Parthenon is built to the divine proportions:its width is 1.618 times its height. Besides divine proportion, basic principles of proportion are also used. For example, the Pentagon is made up of five internal (or concentric) pentagons. Each of these internal pentagons is in proportion to the outer pentagon.

The concept of proportion is used widely in modern architecture as well. Apartment buildings, or houses within the same community may have different sizes of apartments (or houses). However, they are typically in proportion to each other. Sports stadiums also incorporate proportion: the distance between the bases in a baseball field is always proportional to the length (or width) of the field. Similarly, the width of a goal post in a soccer field is proportional to the width of the entire field.

In addition, architects design miniature models before building the actual structure. These models, known as scale models, serve as detailed representations of the final structure. These scale models are much smaller in size, but are in proportion to the final structure. For example, if the scale model of a house is a hundred times smaller than the actual house, every room (or part) of the model would also be a hundred times smaller than the corresponding room (or part) in the actual house—all parts of the model are in proportion with the actual house. Similarly, different parts within the model are also in proportion. If the actual house should be built such that there are two rooms—one room twice the size of the other, the model would also depict two rooms, where the size of one room is twice that of the other. Simply put, the ratio of the sizes of the two rooms is equal in both cases.

The main advantage of a model is that it allows the architect to visualize a structure before it is built. Also, once the model is created, using proportion, various measurements of the final structure can be easily determined and constructed accordingly.



The proportions of man are carefully delineated in the drawing *Vitruvian Man* by Leonardo da Vinci. CORBIS CORPORATION. REPRODUCED BY PERMISSION.

ART, SCULPTURE, AND DESIGN

Like architecture, painting and sculpting also relies on the concept of proportion. Some of the great painters and sculptors, for centuries, have used mathematical models of proportion to attain visual appeal and symmetry (balance) in their work. Portraits and paintings depicting natural scenery are, more often than not, in proportion with the real thing. For a portrait of a person, a good painter would ensure that the measurements of body parts in the painting are in proportion to the actual measurements of the person. This can be seen in most of the ancient as well as modern day portraits.

In addition, different elements within the same painting are also in proportion. In a painting of natural scenery depicting a house, trees, fences, and mountains, the size of each of these is not similar. A house in the painting would be bigger than the size of the fence (unless they are supposed to be at different locations far away). In other words, depending on their location, the sizes are always in proportion—similar to what we see in the real world.

The same holds true for sculptures as well. Like a painting, the sculpture of a person may be bigger (or smaller) in size than the person. However, in most cases, the measurements are in proportion. The advantage of proportion for creating sculptures is evident when the difference in size of the actual object and that of the sculpture is large. Mount Rushmore, in South Dakota, is a classic example. The design and development of the famous memorial to the four presidents—George Washington, Thomas Jefferson, Abraham Lincoln, and Theodore Roosevelt, is based on a number of mathematical concepts, such as ratio, proportion, and scale.

Prior to sculpting the faces of the presidents on the mountain itself, the designer of the memorial, Gutzon

Borglum, developed a smaller model. The size and measurements of the memorial on the mountain are in proportion to the model. Carving the faces on the mountain directly would have been an extremely difficult task for the designer and his team. However, a smaller but proportional model greatly simplified the process. Many technical aspects such as distance between the faces, size of each face, measurements within a particular face, could be easily calculated in the model. Once all measurements were recorded, the designer used the proportion equation to calculate the actual measurements of the memorial (in order for it to look exactly like the model itself).

The principles of interior design also rely on proportion. Furniture, for example, is designed so that its parts are proportionate to each other. This is critical in achieving stability and balance. Furniture that is out-of-proportion is not considered visually appealing. The parts of a chair—the arms, legs, seat, and back—are in proportion to each other, and the chair as a whole.

MEDICINE

Medicines are very essential to treat many illnesses and diseases. Medicines are also used during surgeries and medical diagnosis. They often contain more than two ingredients or compositions that are essential to have desired effect. The proportion of this composition becomes very important. In other words, every medicine contains a specific proportion of its ingredients.

Prescriptions, as well as over the counter drugs, require the mixture of various chemicals, and other additional constituents, to be in certain proportions. For example, over the counter medicines for pain relief often contain aspirin, a required drug to relieve pain, along with other drugs. The proportion of each constituent present in medicine is important as they are meant to treat a certain type of disease, illness, or ailment.

Changing the proportion of the constituents can have different effects. Several common ingredients are used to treat different types of illnesses. The reason for this is that medicines, when prepared using different proportions of the same drugs (or ingredients), act differently, and hence are meant for different diseases.

Proportion is also used frequently by doctors and nurses, while preparing dosages for patients. Patients may require dosages of drugs that vary in quantity and strength. For example, some times a patient may need a dosage that contains 200 mg of a drug that comes as 100 mg diluted in 1 ml of fluid. The technical specifications associated with dosage measurement are beyond the scope of this article. However, for our purpose, the above dosage can be thought of containing a drug in specific quantity (200 mg), having specific strength (100 mg diluted in 1 ml of fluid). The quantity of a drug is proportional to its strength. Using this relation, health care professionals can calculate the quantity of the drug to be administered for a particular strength.

MAPS

Maps may represent a large geographical area and can be of various types depending on the features they emphasize. The area represented by a map can vary from a small room to the entire universe.

There exists a relationship between a specific distance on the map and its actual distance. This relationship is defined by the mathematical concept of scale (or map scale). However, it is important to note that the map scale is based on proportion. In simple terms, the size of the map and the size of the area it shows are always in proportion.

Consider a map that depicts an area that is a hundred times larger than the size of the map. In this case, the relationship between the map and the actual area can be shown as the map scale (a ratio in this case) 1:100—one unit of measurement (cm, inch, feet, etc.) on the map is equal to hundred units in the actual area. The ratio between any part of the map to its actual size remains the constant (1:100). Therefore, every part of the map is in proportion to its actual size. For example, if the actual distance between two points is 100 inches, then the distance between the same two points on the map would be 1 inch. Similarly, if the actual distance between two points is 500 inches, the distance between these two points on the map is 5 inches—the distance between any two points on the map is proportional to the actual distance between them.

Maps can be categorized into two types—the large scale map, and the small scale map. The large scale map shows a smaller area but in greater detail, whereas a small scale map shows a larger area in less detail. The map scale for these maps would differ; however, the maps are always in proportion to the actual size. A city map would be an example of a large scale map as compared to a world map (small scale).

ERGONOMICS

Ergonomics is a science that studies technology and how well it suits the human body. Ergonomics involves understanding basic body parts, their functions and abilities to operate equipments, machinery, products, and other technological devices. Ergonomics is commonly used while designing cars, among other things. Ergonomic car designs are based on the principles of proportion.

Consider, for example, a car seat for drivers. Its height from the surface, inclination, and movements patterns are all designed in proportion to the human body. The size of the seat has to be in proportion with the size of an average human driver. In addition, you do not expect a person to have a giant steering wheel in front of him/her—the size of the wheel (the diameter of the wheel) has to be in proportion to the size of the hand grip, shoulder width, and distance between the wheel and person driving the car.

Ergonomics is used extensively in many areas as well. This includes design of kitchen and appliances, design of home and office furniture, bathroom appliances, electronics, computer systems, airplane and train interiors, and much more. Every ergonomically designed object is proportional to the size of the human body (or a part of it).

For example, a bed is usually designed in proportion to the human body. The length of a bed is proportional to the average height of a person. Many beds in Europe are around seven feet (2 m), whereas those in Asia are around six feet (2 m) long. This also influences other design standards such as height of the bed from the floor, and width of the bed.

ENGINEERING DESIGN

Engineers apply the principle of proportion in many ways including when designing automobiles, airplanes, and trains. Representative two-dimensional models (similar to scale models discussed earlier) are designed before finalizing and manufacturing a car, plane, or train. These are detailed models depicting each and every characteristic. The automobile is then built such that its size and other measurements are directly proportional to the model. In other words, a relation based on proportion is established between the model and the actual object.

The main benefit of creating models for automobiles (as well as airplanes and trains) is to easily study design issues. For example, after calculating the measurements of a seat in the car model, using proportion, the actual size of the seat can be calculated. This will enable the designer to analyze whether the size of the seat is appropriate for a person.

As the dimensions and size of the car are proportional to the model, any change in the model would affect the car. Besides, parts of the model (or car) are also proportional to the model (or car) as a whole. Put simply, if for example, the size of the leg room is changed, the change in the total size of the car can be calculated. if leg room needs to be increased, and at the same time the size of the car must remain constant, the designer would have to reduce the size of some other part of the car.

Once a model with ideal measurements is created, manufacturing the final object becomes a lot easier.

MUSICAL INSTRUMENTS

Since ancient times, mathematicians have always established relationships between principles of mathematics and music. Pythagoras was the first people known to study and apply concepts of proportion and scale to music. These principles are also valid for most musical instruments.

It is widely believed that instruments designed using specific proportions produce superior music. This can be seen in both ancient as well as modern day instruments. For example, to achieve better quality of music, the distance between strings on a guitar (or a violin) is proportional to its entire width. In fact, proportion is used for designing every part of the instrument. Similarly, for a piano to function properly, all its parts have to be in proportion to one another.

CHEMISTRY

Chemicals are often a mixture of a variety of substances. These substances are present in certain ratios. For example, the chemical composition of ammonia is NH_3. Here, the amount of nitrogen (N) is directly proportional to the amount of hydrogen (H)—the ratio of nitrogen atoms to hydrogen atoms is 1:3. In other words, if the number of nitrogen atoms increases by one, the number of hydrogen atoms have to be increased by three. Similarly, if two nitrogen atoms are added, six hydrogen atoms must also be added to continue for the substance to be ammonia. The ratio between nitrogen and hydrogen is always maintained.

Setting up equations as proportions is one of the most effective ways of solving a number of problems in chemistry. For example, to prepare chemical solutions, the chemicals are usually dissolved in water or alcohol. The quantity of chemical present in the solution is known as the strength of the solution. In simple terms, a 70% solution would contain 70% of chemical and 30% of alcohol (or water). While preparing the solution of a specific concentration, the amount of chemical is always proportional to the amount of alcohol (or water). This relationship is especially useful while preparing solutions in different quantities but the same concentration.

A 50 mL (four tablespoons) of chemical solution contains 20 mL (a little more than one tablespoon) of alcohol. If the amount of chemical solution has to be increased to 80 mL (a little more than five tablespoons), what would be the amount of alcohol present in this solution? This

Key Terms

Ratio: The ratio of a to b is a way to convey the idea of relative magnitude of two amounts. Thus if the number a is always twice the number b, we can say that the ratio of a to b is "2 to 1." This ratio is sometimes written 2:1. Today, however, it is more common to write a ratio as a fraction, in this case 2/1.

Proportion: Two quantities with equal ratios.

Scale: The ratio of the size of an object to the size of its representation.

Symmetry, or balance: A design is symmetrical if its two opposite sides divided by a line in the center are identical, or nearly identical.

can be calculated by setting up a proportionality equation as shown below:

20 mL alcohol / 50 mL solution = x mL alcohol/ 80 mL solution, where x is unknown amount of alcohol. The quantity of alcohol should be 32 mL (two tablespoons) for an 80 mL solution.

Such equations are used widely by doctors, scientists, and students.

DIETS

Dieticians and fitness experts often apply mathematical approaches to developing "balanced" diets. They indicate that every meal should have proteins, carbohydrates, and fats in a certain proportion to each other (and the entire meal). This relationship helps greatly in calculating the amount of proteins, carbs, and fats for different meal portions.

For example, a particular meal amounts to 400 calories—160 calories from proteins, 160 calories from carbohydrates, and 80 calories from fat. If another meal is equivalent to 600 calories, the amount of proteins, carbs, and fats would increase to 240 calories, 240 calories, and 120 calories respectively. Note that the amount of proteins, carbs, and fats is in proportion.

Most food items list the amount (in grams) of protein, carbohydrate, and fat content. For instance, 100 grams (3.5 oz) of ice-cream may contain 20 grams (0.7 oz) of fat. The amount of fat in 50 grams (1.7 oz) of the same ice-cream would be 10 grams (0.3 oz), and so on—fat content is proportional to the total quantity. Food items are always available in specific quantities. Put simply, by applying proportion equations, the content of proteins, carbohydrates, and fats can easily estimated for different quantities.

The same concept is also applied to cooking. While preparing a food item, the ingredients are in proportion to each other (and to the total quantity of the food item).

STOCK MARKET

Mathematical concepts such as proportion and ratio have a lot of business applications. One such example is in the stock market. There are factors that contribute to the share value of a company. However, more often than not, a company's share value fluctuates based on profit it makes. Besides, the value also depends on the number of buyers of the company shares. Simply put, the value of a share is proportional to a combination of factors, including the profit and number of buyers.

Most companies divide a percentage of profits amongst all its shareholders (people who own the company's shares). The amount given per share is known as dividend. Higher the number of shares a person owns, higher the dividend. Another way to look at this is that the total dividend is proportional to the number of shares owned.

PROPORTION IN NATURE

The number Phi is an unusual number with astounding mathematical properties. As explained earlier, the golden section, a principle on which ancient Greek architecture was based, is derived from a ratio that further results in the number phi. Phi appears in proportions of the human body as well as the proportions of various other animals. The renaissance artists referred to the golden section as the divine proportion and used it for achieving balance in arts. The divine proportion principle is found in abundance in nature. The spirals of a sea shell, the galaxy, the body of a dolphin, the structure of a butterfly, a peacock feather, the patterns of flowers and plants, the rings of Saturn, all follow the divine proportion principle.

The average human face is also an example of divine proportion. The head forms the golden rectangle with eyes exactly at the center. The mouth and nose are each placed at golden sections of the distance between the eyes

and the bottom of the chin. Assume that the eyes are represented by A, nose by B, mouth by C and chin by D. The ratio of line AC to line AD is the same as ratio of line BC to line AC. This means that the ratio of distance between eyes and mouth to the distance between eyes and chin is in proportion with the ratio of distance between nose and mouth and eyes and mouth. Some scientists who study psychological reactions to faces assert that concepts of beauty may be related to facial symmetry and proportion.

Interestingly, the average human face, when viewed from side also reflects the divine proportion principle. Even the dimensions of human teeth are based on this principle. Some dentists are even considering the knowledge of this principle to enhance their aesthetic dentistry skills. The human hand is also an example of the divine proportion.

Where to Learn More

Books

Elam, Kimberly. *Geometry of Design: Studies in Proportion and Composition.* New York: Princeton Architectural Press, 2001.

Padovan, Richard. *Proportion: Science, Philosophy, Architecture.* London: E & FN Spon, 1999.

Quadratic, Cubic, and Quartic Equations

Overview

An equation often describes a function, a rule that relates numbers in one set to numbers in another. Rather than listing all the numbers related by a function, letters, also termed variables, are often used to stand in for the numbers.

Fundamental Mathematical Concepts and Terms

The function $y = 2x$ says that for every number x in some set there is some other number, y, in some other set that is twice as large as x. Some functions consist of a sum of powers of x, like $y = x^3 + 3x^2 + 2x + 1$.

Here the number just above each x tells us how many times to multiply x times itself: that is, $x^3 = x \times x \times x$, and so forth. Functions of this form are named by the highest power of x they contain, which is the rank or order of the equation. For example, the highest power of x in $y = 2x$ is 1 (because $x = x^1$), so this is a first-order equation. The highest power of x in $y = x^3 + 3x^2 + 2x + 1$ is 3, so this is a third-order equation.

The first four orders have special names, namely linear, quadratic, cubic, and quartic. Quadratic and higher-order equations appear constantly in science, engineering, and business mathematics. They are used literally millions of times a day in these fields, designing electronics, analyzing data, implementing codes, predicting profits, and performing other tasks.

Examples of equations of the first four orders are given in Table 1. In the examples, the letters A through E are used to stand for any constants (fixed numbers), with the exception that A cannot equal 0. These constants are called the coefficients of the equation.

A "solution" to an equation is an x, y pair for which the equation holds true. For example, a solution to the linear equation $y = 2x$ is $x = 5$, $y = 10$, because $10 = 2 \times 5$. In this equation—in fact, in all linear equations—there is one x for each y. Finding solutions to equations is one of the most common tasks in the mathematics of science, engineering, and business. Often we know what y is, or what we want it to be—the cost of an item to be manufactured, say—and we want to know what x (or x's) will produce that y. The variable x often stands for something that we can chose or control, such as the length of an assembly line or the amount of a chemical added to a reaction.

For equations where y is equal to a sum of powers of x, including linear, quadratic, cubic, and quartic equations, the x's for which the equation is true are called its

roots. Often the y value is subtracted from both sides of the equation to produce a nice, neat 0 on the left-hand side of the equation, but this is a minor detail. What is important is that the number of roots is equal to the order of the equation. A linear (first-order) equation has one root, a quadratic (second-order) equation has two roots, and so on.

We can find the roots of any linear, quadratic, cubic, or quartic equation by writing down certain equations containing the coefficients of the original equation. This cannot be done for equations of order higher than 4, as mathematicians have known since the 1820s. The first four orders are therefore special. The equation that gives the roots of a quadratic equation, $y = Ax^2 + Bx + C$, is one of the most commonly used formulas in all math and science, and has been known since mathematicians of Babylon discovered it some 4,000 years ago:

$$x = \frac{-B \pm \sqrt{B^2 - 4AC}}{2A}$$

This formula is known as "the quadratic equation." In the equation $0 = 2x^2 + 3x - 1$, we have $A = 2$, $B = 3$, and $C = -1$ and the quadratic equation gives us the two roots:

$$x_1 = \frac{-3 + \sqrt{3^2 - 4 \times 2(-1)}}{2 \times 2} = .28 \qquad x_2 = \frac{-3 - \sqrt{3^2 - 4 \times 2}}{2 \times 2}$$

These roots are the two values of x for which $0 = 2x^2 + 3x - 1$ is true. If you plug either of them in for x and do the arithmetic on a calculator, you'll see that 0 really is the answer. (The small numbers hanging off x_1 and x_2 are just labels to set them apart.)

Type of equation	General form	Example
Linear	$y = Ax + B$	$y = x + 10$
Quadratic	$y = Ax^2$	$y = 2x^2 + 3 - 1$
Cubic	$y = Ax^3 + Bx^2 + Cx + D$	$y = 12x^3 + x + 5$
Quartic	$y = Ax^4 + Bx^3 + Cx^2 + Dx + E$	$y = x^4 - 12x^3 + x^2 + 100$

Table 1.

Real-life Applications

AREA AND VOLUME

The most basic uses of quadratic and cubic equations are for determining area and volume. In fact, it was the need to calculate land areas that motivated the Babylonians to discover the quadratic equation to begin with. You already know that the area of a square with edges x units long is $x \times x$ or x^2. If we call the area of a square S, then we have the quadratic equation $S = x^2$ (which can also be written $0 = x^2 - S$). The formulas for the area of a circle, a triangle, or even of the surface areas of solids like spheres and cubes, all contain x^2; all are quadratic equations. Surface area is important in real estate, medicine, physics, and engineering. It affects how fast an object cools off (greater area equals quicker cooling), which is why machines that need to get rid of extra heat sometimes have little metal fins stuck on them to increase their surface area. It affects how quickly a droplet evaporates (greater area equals quicker evaporation). It affects how quickly a chemical reaction proceeds (greater area equals quicker reaction).

Cubic equations come up just as naturally. Recall that the volume of a cube with an edges x units long is x^3. If we call the volume of the cube V, then we have the quadratic equation $0 = x^3 - V$. And, just as with surface area, this cubic relationship pops up not only in the formula for the volume of a cube, but in the formula for the volume of a sphere or cylinder or any other three-dimensional object.

The fact that area is described by a quadratic equation and volume by a cubic equation affects many things in nature. Any object's surface area is proportional to x^2—where x stands for how wide the object is—but its volume is proportional to x^3. And as you make the object bigger, that is, increase x, x^3 will always grow faster than x^2. This is why insects can't (lucky for us) grow to the size of dogs or whales: they breathe using surface area (x^2) but their need for oxygen goes by body volume (x^3). This is why elephants have fat legs: the strength of a leg-bone goes by cross-sectional area (x^2), but the weight the bone has to bear goes by the volume of the elephant (x^3).

ACCELERATION

Quadratic equations are needed to predict the paths of accelerating objects. Acceleration is any change in speed. When the driver of a car steps on the gas or hits the brakes, the car accelerates (goes faster or slower). When you drop a ball or throw it up in the air it accelerates. And almost any time a machine with moving parts is designed, from a CD player to a car engine to a jet plane, the people designing the product must deal with accelerations.

CAR TIRES

Car tires are made of rubber-like plastics derived from petroleum and interwoven with metal wires, and

must work well despite thousands of miles of use, violent blows from bumps, fast turns, and other stresses. Your life depends on them every day, and their design is a complex art. Computer calculations are used to predict how a new tire design will behave, as this is much cheaper than casting actual tires in a trial-and-error way. One of the most important factors in modeling a tire using calculations is describing the mechanical properties of the "rubber" used in the tire: how it responds to stretching, squeezing, and twisting. In a class of new synthetic tire materials called "carbon black filled rubber compounds," it has been found that a cubic equation best describes the stress-strain relationship—that is, how much the material gives in response to a certain amount of force. This cubic equation is used in writing a computer program that will accurately predict how a tire made with these compounds will behave.

JUST IN TIME MANUFACTURING

Traditional economics treated supply and demand as the two factors deciding profitability in manufacturing. However, in the 1990s some Japanese manufacturers introduced a philosophy called "just in time" (JIT) manufacturing. In this approach, a manufacturer—say of cars, computers, or cameras—tries to produce as many items as possible just in time to deliver them to a buyer. Manufacturing a product and then having it sit in a warehouse, waiting to be sold, reduces profit. But a manufacturer must balance certain variables: they must announce a price and stick to it, they must guess at how much delay or "lead time" they will need to deliver a product, and they must guess at how much demand for the product there will be. The goal, as always, is to earn maximum profit. It turns out that the solution of a cubic equation is central to solving the equation for maximizing profit.

HOSPITAL SIZE

Since the 1980s, hospitals have found it increasingly difficult to make a profit—or even to stay out of debt. Mathematical cost-profit analysis has therefore been brought into play to help hospitals make more profit. One basic decision that a hospital must make is how many beds to have. Having too few or too many beds makes it harder for a hospital to be profitable. Traditionally, profitability has been described as a quadratic function of bed size (the number of beds in the hospital, not how big each bed is); more recent work has shown that a cubic equation works even better. (Other factors are involved, such as where the hospital is located and how affluent the surrounding population is. But if these assumptions are held steady, profitability is a cubic function of bed size.) Using a cubic equation, researchers have found that there isn't just one bed size that is most profitable, but two; or, rather, a point this is typical of a cubic equation, which can have two maximum points rather than one (as a quadratic equation does). From 0 to 238 beds, profit increases. After 238 beds it decreases until 560, after which it goes up indefinitely (but other factors prevent us from building infinitely large hospitals). A hospital is therefore most profitable, in the United States under current conditions, if it is either medium-sized (about 238 beds) or as big as it can be (560 beds or larger).

GUIDING WEAPONS

In steering weapons such as missiles and planes, it is necessary to tell the computer that guides the weapon where it is. Each position is coded as a set of numbers, the "coordinates" of the weapon or vehicle. These can be given in traditional terms as latitude and longitude (numbers derived from a network of imaginary lines laid down on the Earth's surface by map-makers) plus altitude (height above the surface), or in terms of an "Earth-centered coordinate system." Since one type of coordinates is better for some purposes and the other is better for other purposes, it is sometimes necessary to translate between them—to take position information given in one form and turn it into the other form. Going from latitude-longitude coordinates to Earth-centered coordinates is mathematically easy, but going the other way requires the solution of a quartic equation.

Where to Learn More

Web sites

Budd, Chris, and Chris Sangwin. "101 Uses of a Quadratic Equation." *Plus Magazine*. March 29, 2004 and May 30, 2004. Part I: <http://plus.maths.org/issue29/features/quadratic/index-gifd.html>. Part II: <http://plus.maths.org/issue30/features/quadratic/index-gifd.html> (Oct. 22, 2004).

Ratio

Overview

A ratio defines the numerical relationship between two comparable quantities. Examining the ratios between two or more values often provides valuable insight into the patterns and behaviors of numbers.

Ratios exist naturally throughout the universe. The ratio of the size of one planet to another nearby planet can affect the orbits of both planets. The ratio of owls to mice plays a big role in the survival of both species. The ratio of height to trunk width limits the growth of trees. Humans have used ratios in almost all of our creations throughout history. The physical stability of a building depends on several ratios—involving height, width, angles, and the strength of materials that must be carefully analyzed to ensure the safety of the people inside. The accurate mixing of chemicals that allows us to create stronger materials is also reliant on ratios that define how much of each substance is needed with respect to the other materials. People around the world use ratios on a daily basis to organize time and finances.

Fundamental Mathematical Concepts and Terms

A ratio between two numbers X and Y is usually expressed in one of three ways:

- X/Y (much like a fraction)
- X:Y
- "X to Y"

Each of these expressions represents the ratio of X to Y.

For example, if there are 12 cars for every three trucks, then the ratio of cars to trucks can be written as 12/3, as 12:3, or as "12 to 3." Given this information about cars and trucks, it is also true that the ratio of trucks to cars is 3/12, 3:12, or "3 to 12".

All of these expressions for the ratio of cars to trucks (or trucks to cars) state exactly the same thing: for every 12 cars, there are three trucks. Suppose that people in a certain neighborhood always keep their cars in their garages, but leave their trucks out in the driveway. If three trucks are visible in the neighborhood, then there are 12 cars in the neighborhood, even though they are hidden in garages.

The foundation of the idea of a ratio is that whatever happens to one of the numbers also happens to the other. Suppose that six trucks can be seen in driveways around the neighborhood. This means that there are 24 cars hidden in garages. The number of trucks was doubled (multiplied by 2) so the number of cars must have doubled as

well. Division of ratios works in the same way. If there was only one truck in the entire neighborhood, then there would be only four cars. Here, the number of trucks and cars are both divided by two to arrive at the ratio 1:4. In fact, this is the simplest form of the ratio of trucks to cars. In a case such as this, the ratio can be simplified so that one of the values is one, which is a good illustration of how ratios work: no matter how many trucks are in the neighborhood, the number of cars is four times as large. Not all ratios can be simplified this neatly—2:3 for example. In cases like this, a decimal can be used as 2:3 simplifies to 1:1.5. In any case, it is easiest to understand the relationship between the two values when the ratio is simplified.

Ratios can be multiplied together to discover new ratios. For instance, if there are two cars for every truck, and three trucks at every house, then there are six cars at every house. That is, 2:1 multiplied by 3:1 is equal to 6:1. Perhaps money provides a better illustration of this concept. There are four quarters to every dollar and five nickels to every quarter; so there are 20 nickels to every dollar. This can be verified by multiplying the five pennies in each nickel by 20 (the number of nickels in a dollar) to get 100 pennies to every dollar.

Although often expressed as a quotient (one number divided by another, such as 2/3), ratios are not the same thing as fractions. For example, if Otis has two dogs and four cats, then the ratio of dogs to cats in his house is two to four, which simplifies to 1:2 or 1/2. This indicates how many dogs there are compared to cats (there are half as many dogs as cats). However, the fraction of animals in Otis' house that are dogs is two out of the total number of animals or 2/6, which simplifies to 1/3. This means that one third of all of his animals are dogs. Be careful to understand how fractions are related to ratios when using the quotient style of notation. To avoid confusion, this text most often uses the X:Y style of notation for ratios.

A Brief History of Discovery and Development

The term ratio stems from an early sixteenth century Latin word meaning reason or computation. However, the mathematical concept of ratios helped people understand the universe around them long before that.

For example, the relationship between a circle's diameter (the length of any line connecting one side of the circle to the other through the center of the circle) and circumference (the length of the boundary of the circle) was approximated for thousands of years before the Greek mathematician Archimedes discovered a way to define the relationship exactly. This ratio can be used to determine the circumference of a circle if its diameter is known, and vice versa. The circumference of any circle is equal to the diameter multiplied by this ratio, commonly represented by the Greek letter pi, and approximately equal to 3.14159265.

Ancient Egyptians approximated pi (though they did not call it pi) as 3.1605. The Old Testament of the Judeo-Christian Bible contains a reference to an approximation of 3:1 for the ratio of a circle's radius to the circumference of a circle. Although ancient Babylonians generally agreed with this approximation throughout most of their history, a stone tablet believed to have been created by Babylonians sometime between 1900 and 1680 B.C. referred to a slightly more accurate approximation of 3.125 for pi.

Early approximations of pi were dependent on approximations of the circumference of circles. It is believed that most approximations of circumference were found using methods similar to those used by Archimedes. First a circle was placed inside of the smallest hexagon (a polygon with six sides) that it could fit into. The length of the perimeter of the hexagon was calculated by measuring one side and multiplying this value by six. Next, the perimeter of largest hexagon that could fit inside the circle was calculated. Because the smaller hexagon just barely fits into the circle, and the circle just barely fits into the larger hexagon, the circumference of the circle is somewhere between the lengths of the perimeters of the two hexagons. To arrive at a better approximation, the number of sides of the two surrounding polygons was increased. As more sides were added, the two polygons fit the circle more snugly and the perimeters became closer and closer to the circumference of the circle. Archimedes used these approximations as clues that eventually led him to find a way to define the ratio of diameter to circumference exactly.

Another important ratio studied throughout history is the Golden Ratio, also known as the Golden Mean, the Divine Section, the Golden Section, the Golden Cut, the Divine Proportion, and many other names. The main reason that this ratio has so many names is that it has been discovered at different times by civilizations that use different languages and, most importantly, different numbering systems. The Golden Ratio is approximately 1.6180339887498948482 to 1 (how the Golden Ratio is calculated is beyond the scope of this text). The Golden Ratio is usually denoted by the Greek letter *phi* (ϕ).

The Golden Ratio can be found throughout nature—from the patterns found in leaves, pinecones, and seashells, to the reproductive patterns of certain animal

species. It is also argued that the Golden Ratio provided a basis for the architecture of the ancient Egyptians (including the designs of pyramids and tombs), Greeks (the Parthenon), and Romans. Some ancient Egyptian hieroglyphics show signs of the Golden Ratio as well. Leonardo da Vinci, Mozart, and Beethoven purposely incorporated this ratio into their works. The seemingly endless applications of the Golden Ratio provide brilliant illustrations of the fascinating relationships between numbers.

Real-life Applications

LENGTH OF A TRIP

Ratios can be used to estimate length. For an example let us assume that Tom needs to drive from New York to Miami for a business convention on Saturday evening. He has never driven that far and wants to figure out about how long it will take, so he buys a map of the United States. He notices two bars labeled Scale in the corner of the map. The longer of the two bars represents 100 miles, and the shorter bar represents 100 kilometers. He uses his ruler and finds that the 100-mile bar is one inch long; so the ratio of inches to miles on Tom's map is one to 100. Using the other side of his ruler, he finds that the 100 kilometer bar is one centimeter long; so the ratio of centimeters to kilometers is also one to 100.

Tom is more comfortable thinking in terms of miles, so he chooses to approximate the length his trip based on the inch to mile ratio of 1:100. All he needs to do is find out how many inches separate New York and Miami on the map. Tom lays his ruler on the map, with the beginning of the ruler (representing zero in inches) at New York. The shortest driving route is not a straight line, so he must approximate how long, in inches, his route is on the map. Starting from New York, he measures one inch in the direction of the route that he will take, and marks the spot on the map with a pencil. Then he moves the beginning of the ruler to the mark he just made and measures another inch, following his intended route as accurately as possible. Continuing in this way, he makes 13 marks. The last mark is a little past Miami on the map, so he figures that the route is a little less than 13 inches long. He can't be late to his convention, so he decides to use 13 inches as the base of his calculations. As he found before, the ratio of inches to miles represented on the map is 1:100.

Tom then wants to figure out how many miles are represented by 13 inches, so he must multiply the ratio through by 13 to get a ratio of 13:1,300. This ratio indicates that 13 inches on the map represents 1,300 miles in the real world. So Tom's trip will be about 1,300 miles in distance (length).

Tom now needs to utilize another ratio to help him decide when to leave New York. Without exceeding the speed limit, he can drive about 500 miles in a day. So his mile to day ratio is 500:1. This means that he can drive 500 miles in a single day, 1,000 miles in two days, 1,500 miles in three days, and so on. He needs to go a total of 1,300 miles, so he cannot make it in two days. He can make it easily in three days. He may be a little early but he will not be rushed. He decides that if he leaves on Thursday morning, he will get to the convention with time to spare.

COST OF GAS

In the previous example, Tom calculated 1,300 miles as a slight overestimate for the length of his trip from New York to Miami. He now wants to calculate how much money he will need for gas so that he can plan the budget for his trip. His car gets an average of 25 miles per gallon, which is a mile to gallon ratio of 25:1. Tom uses this ratio to calculate how many gallons of gas his car will need to go 1,300 miles. 1,300 miles is 52 times as long as 25 miles, which means that Tom must multiply both sides of the ratio by 52. In this way, he calculates the mile to gallon ratio 1,300:52. To go 1,300 miles, Tom's car will need 52 gallons of gas.

Next, Tom looks on the Internet and discovers that the average cost of gas along his route is two dollars per gallon. So the ratio of dollars to gallons of gas is 2:1. To find out how much 52 gallons of gas will cost, Tom multiplies both sides of the ratio by 52 to get a dollars to gallon of gas ratio of 104:52, meaning that Tom needs $104 to buy 52 gallons of gas for his car. After working this figure into his budget, he finds that he has plenty of money for his trip to Miami.

GENETIC TRAITS

In 1866, Austrian monk and geneticist Gregor Johann Mendel (1822–1937), published his results from an extensive series of experiments that investigated how characteristics are passed to offspring. One such experiment involved the cross-pollination (transferring the pollen of one plant to another) of two different varieties of pea plants, a green wrinkly pea plant and a yellow rounded pea plant. In this experiment, Mendel discovered that the ratio of yellow rounded offspring to green wrinkly offspring was 3:1, meaning that the cross-pollination process produced three yellow rounded pea

China's one-child family planning program, in combination with a preference for male children, has created an unbalanced boy-girl ratio according to U.S. State Department officials. AP/WIDE WORLD PHOTOS. REPRODUCED BY PERMISSION.

plants for every green wrinkly pea plant. This suggested that the yellow characteristic is three times as likely to appear in the offspring as the green characteristic, and the round characteristic is three times as likely to appear as the wrinkly characteristic. This dominance of yellow and rounded characteristics led Mendel to believe that there are two different types of genetic traits: dominant traits and recessive traits. In these two types of pea plants, the yellow color and round shape dominate the green color and wrinkly shape.

Mendel used this idea of dominance to explain why the ratio of dominant traits to recessive traits is always 3:1. For instance, if the dominant yellow color trait is represented by Y, and the recessive green color trait is represented by g, then the four possible combinations of these traits are YY, Yg, gY, and gg (where each parent plant provides either a Y or a g). A dominant trait only needs to appear once in the combination in order for the dominant characteristic to appear in the offspring. Y appears in three of the combinations, and the only combination that results in a green pea plant is gg. Therefore, the ratio of offspring that show a dominant characteristic to offspring that show a recessive characteristic is three to one, or 3:1. This is true for the shape trait as well. Keep in mind that Mendel's actual experiments and results were more complicated than described here.

Mendel's experiments and conclusions explained a phenomenon that had confused people for thousands of years, that traits can appear after skipping generations. For example, cross-pollination of two yellow pea plants can result in a green offspring as long as at least one parent had a Yg or gY gene.

The importance of Mendel's results was not truly recognized until the beginning of the twentieth century when multiple researchers independently rediscovered Mendel's conclusions in their own experiments. Since then, Mendel's findings have been the foundation of many genetic studies and practices, including the creation of new flowers and new species of pet fish, and enhancements in

farm production that improve the quality of produce found in most grocery stores.

STUDENT-TEACHER RATIO

The student-teacher ratio compares the number of students to the number of teachers at a given school. For example, if a school has 44 teachers and 968 students, the student-teacher ratio at this school is 968:44, which simplifies to 22:1. This can be interpreted in a few different ways. A mother might see it as an indication of how much attention her child will receive in school (e.g., her child will share each teacher with 22 other students). This ratio enables a teacher to predict how many students he or she will teach and how many papers she will be grading (about 22 students per class multiplied by the number of classes he or she teaches). A school official may see it as an indication of how many teachers need to be hired in the following year. A prospective college student will usually take the student-teacher ratio into consideration when choosing a school for higher studies.

MUSIC

Ratios can be found in every facet of music. Rhythm, the speed and pattern of beats, has been the foundation of music dating back to ancient drum circles. Rhythm determines how many beats there are in a measure (the standard unit in the arrangement of song). For example, if there are four beats in a measure, then the ratio of beats to measures is 4:1. This means that the musician considers four beats—possibly counted by tapping a foot—to be a single standard unit in the arrangement of the song. Different ratios of beats to measures affect different types of music. For example, a typical waltz has a ratio of 3:1.

The relationship between the pitch of two musical notes is also a ratio. Whether created by your vocal chords, a guitar, or a finger moving around the rim of a wine glass, the sound of a note is determined by the frequency (speed) of the vibration causing it. As you move from left to right on the keys of a piano, the difference from one note to the next is determined by the ratio between the frequencies of the two notes: the ratio between the frequencies of two subsequent notes is always the same. Harmony (whether or not two or more notes sound good when played together) is determined by the ratios of their frequencies as well.

Ratios in music allow songwriters and musicians to communicate the intended shape and feel of a song. The many ratios in a composition define how the various sounds relate to each other in time and, whether consciously noticed or not, give the music both structure and beauty.

AUTOMOBILE PERFORMANCE

The safe and efficient operation of any automobile is dependent on many ratios. Oils and fluids must be present in certain ratios to keep the engine and brakes operating properly. The relationships between the size, weight, and position of various parts ensure that a car or truck can make turns while traveling at reasonable speeds and can stop quickly when necessary. Two ratios found in automobiles are compression ratios and gear ratios.

The compression ratio is used to predict how efficiently an engine will perform. In general, a higher compression ratio indicates better engine performance. High compression ratios are often associated with requirements of more expensive fuel and frequent engine maintenance. The determination of an engine's compression ratio involves the relationship between the sizes of the parts of the engine that cause combustion (the small explosions that provide an engine with power).

The speed and power of an automobile depend partially on the ratio between the sizes of gears that cause the wheels to turn. A larger gear turns slower because it has more teeth and takes longer to complete a full revolution. If a gear that is powered by the engine is attached to a smaller gear, the smaller gear will turn more quickly than the large gear. This increases the speed of revolution without increasing the need for power. Given certain gear ratios for an automobile, a specialist can determine how many revolutions per minute (RPM) are required to go a certain speed, or how many tons can be pulled without overexerting the engine. A typical car has multiple sets of gears intended to perform different actions. The first gear has a high gear ratio in order to provide the car with enough power to get the car started. In higher gears, the gear ratio is increased in order to enable faster speeds. Also, a car would eventually get stuck without an additional gear set that caused the car to move in reverse.

SPORTS

Ratios are often used to assess the performance of an athlete or athletic team. The relationship between two or more statistics often proves a better indication of performance than a single statistic alone.

As an example, a point guard's contribution to a basketball team is partly measured by his assist-to-turnover ratio. This ratio is determined by comparing the number of assists (passes that lead to an immediate basket) to the number of turnovers (anything that causes the ball to be lost to the other team). Suppose Gary has had 53 assists this year, and has turned the ball over to the other team 44 times. Gary's assist-to-turnover ratio is 53:44.

Gary's talents could be judge based on turnovers alone. If Gary had more turnovers than anyone else this season, sports analysts might think that he is the worst point guard because he gives the ball up more often than anyone else. But what if he also happened to have the most assists? Would the analysts still think so poorly of him? The converse is true as well: if Gary's talents were judged based only on the number of assists that he has without taking into account the fact that he turns the ball over quite often, the analysts would not have a very accurate picture of how Gary actually performs on the court.

AGE OF EARTH

In 1905, New Zealand/English physicist Ernest Rutherford (1871–1937) announced a discovery that would forever change the approximations of the age of Earth. He suggested that the age of rocks could be computed by analyzing one of two ratios: the ratio of uranium to lead or the ratio of uranium to helium. These ratios can be used to determine how long radioactive materials have been decaying, and in turn, to determine how long ago rocks were formed. Prior to this discovery, the process of radioactive decay was poorly understood, and guesses at the age of Earth were just that: guesses. Since Rutherford's discoveries, new tools and methods have been derived to improve estimations of Earth's age. For example, calculations in the dating process, including values for decay rates, have been repeatedly improved upon. As of 2005, the best estimation for the age of Earth is in the neighborhood of 4.5 billion years.

HEALTHY LIVING

A person's height-to-weight ratio is the relationship between how tall that person is and how much that person weighs. If a person is six feet tall and weighs 180 pounds, then his height-to-weight ratio is six feet to 180 pounds, or 1 foot per 30 pounds. This ratio can be seen as an indication of how healthy a person is. There are, of course, many other important considerations—including body type, bone thickness, and muscle density—that help determine an individual's optimal weight. All of these factors can be put into terms of ratios.

COOKING

Whenever a chef follows a recipe, he uses ratios to determine how much of each ingredient to stir in. Suppose a chef is cooking his favorite soup for a large dinner party. He has a recipe that tells him how much of everything is required for making enough of the soup to serve 20 people, but there will be 140 people at the dinner party. The ratio of people served by his recipe to the actual number of people that he needs to serve is 20:140, which simplifies to 1:7 (by dividing both sides by 20). This tells the chef that he needs to buy seven times the amount of ingredients suggested by the recipe in order to make enough soup for the dinner party.

The chef can also use ratios to determine how much of one ingredient will be needed based on the required amount of another ingredient. For instance, the chef knows that the ratio of sugar to butter in this recipe is 1:3. This means that the amount of sugar needed to make any amount of this recipe is a third of the amount of butter needed. The chef has already calculated that he needs six cups of butter to make the soup for 140 people. With no further calculations, he knows that he needs two cups of sugar to make this amount of soup.

CLEANING WATER

Chlorine is the main chemical that is used to clean both drinking water and the water in swimming pools. The biggest difference between the processes for cleaning drinking water and swimming water is the concentration of the chemicals, the ratio of the amount of chemicals to the amount of water. This ratio is much lower in drinking water than in swimming water. That is, the water you drink has a smaller amount of chemicals in it than the water in most swimming pools. The concentration of chemicals in drinking water must be precisely monitored in order to ensure that enough chemicals are present to kill bacteria, but not enough to be harmful when swallowed by humans. Water in a swimming pool must contain a higher concentration of chemicals because the water is constantly in contact with contaminants from swimmers and the air above. The fact that a swimming pool is open to the air also allows the chemicals to evaporate, so new chemicals must be added periodically. These ratios between water and chemicals are essential for the different uses of water. Water from a swimming pool is not safe to drink in large quantities; and swimming in water with the concentration of chemicals found in drinking water would quickly result in the growth of algae and bacteria in the pool.

Potential Applications

STEM CELL RESEARCH

Stem cells are special cells in the human body that have the ability to become any type of human cell. This single type of cell can create skin and muscle tissue, bones

Key Terms

Concentration: The ratio of one substance mixed with another substance.

Percent: From Latin for *per centum*, meaning per hundred, a special type of ratio in which the second value is 100; used to represent the amount present with respect to the whole. Expressed as a percentage, the ratio times 100 (e.g., 78/100 = .78 and so .78 × 100 = 78%).

Rate: A comparison of the change in one quantity, such as distance, temperature, weight, or time, to the change in a second quantity of this type. The comparison is often shown as a formula, a ratio, or a fraction, dividing the change in the first quantity by the change in the second quantity. When the changes being compared occur over a measurable period of time, their ratio determines an average rate of change.

Ratio: The ratio of a to b is a way to convey the idea of relative magnitude of two amounts. Thus if the number a is always twice the number b, we can say that the ratio of a to b is "2 to 1." This ratio is sometimes written 2:1. Today, however, it is more common to write a ratio as a fraction, in this case 2/1.

and bone marrow, and organs such as the liver and lungs. This characteristic has made stem cells the main focus of regenerative medicine, a field of research involving the recreation of cells in the human body. The regeneration of cells may be the solution to many problems that have been unsolvable in the past. Potential uses of cell regeneration include regaining skin and muscle tissue lost in physical accidents; allowing someone bound to a wheelchair to walk; and curing diseases such as Parkinson's, cancer, Alzheimer's, and diabetes. Unfortunately, it may be many years before stem cells are regularly used in routine medical procedures.

Scientists have much to learn about manipulating stems cells to create a desired part of the body. Ratios play a big role in many stem cell research projects. For example, the ratio of blood cells in a donor to blood cells in the recipient may be an important factor in the success of stem cell transplants.

OPTIMIZING LIVESTOCK PRODUCTION

In nature, the sex ratio (ratio of males to females) of many species remains close to 1:1 (often referred to as 50:50 or fifty-fifty), meaning that about half of the population is male and about half is female. In species with males that can mate with multiple females, this may not seem a very efficient ratio. Nevertheless, the sex ratio remains approximately 1:1.

Farmers have for millenia artificially kept the ratio of female cattle (cows) to male cattle (bulls) high, because a single male cow can fertilize multiple female cows. Suppose a single male cow can regularly fertilize up to 20 cows. If a dairy farm with 100 cows had 50 males and 50 females, then only half of their cows would be producing milk, and the male cows would not be performing to their capacity. But if there were 5 males and 95 females, then the farm would have more cows producing milk, and the males would be able to do their job at a rate closer to their limit.

DETERMINING THE ORIGIN OF THE MOON

In 2003, German scientists compared the ratios of two elements present in rocks from the Earth, Moon, Mars, and various meteorites to arrive at a better approximation of how and when the moon was formed. The two elements compared were niobium (a metal commonly found in alloy steels) to tantalum (an acid-resistant metal commonly found in dental and surgical instruments).

Most astronomers have long subscribed to the theory that the moon was formed when a celestial body (roughly half the size of Earth) crashed into Earth causing a large mixture of rocky debris from both bodies to fly into space, some of which lumped together to form the moon, while the rest dropped back to Earth. The amount of the Moon that is made up of material from the body that struck Earth has long been passionately debated; as has the amount made up of material from the Earth itself. The percentage of the Moon that is made up of material from the impacting body, for example, was approximated at as low as 1% by some scientists, and as high as 90% by others. By comparing the ratios of niobium and tantalum, the German team of scientists was able to determine that

the amount of the moon that is composed of material from the body that struck Earth is somewhere between 35% and 65%. The rest of the Moon is composed of material from Earth.

The approximate age of the moon, another value that scientists have had a hard time agreeing about, was also refined during these studies. Calculations based on the ratios of niobium and tantalum suggest that the Moon must have been created at about the same time as Earth: about 4.5 billion years ago. As scientists continue to study the moon, the approximations of its composition and age will become increasingly accurate.

Where to Learn More

Books

Livio, Mario. *The Golden Ratio: The Story of Phi, the World's Most Astonishing Number*. Broadway Books, 2003.

Periodicals

Jacobsen, Stein B. "Geochemistry: How Old Is Planet Earth?" *Science* 300, No. 5625 (2003): 1513–1514.

Web sites

National Institutes of Health. "Stem Cell Basics." The official National Institutes of Health resource for stem cell research. June 10, 2004. <http://stemcells.nih.gov/info/basics/basics6.asp> (March 12, 2005).

Rounding

Overview

Rounding is a way of simplifying a number to an approximate value. The intent is to create a number that is easier to envision conceptually and is more practical to use. When a result close to the actual value is sufficient, rounding can be a useful operation.

Fundamental Mathematical Concepts and Terms

WHOLE NUMBERS

Numbers can be rounded to the nearest unit or a larger scale across multiple units, as is appropriate.

The rules of rounding are rudimentary. Irrespective of, for example, the power of ten under consideration, rounding is based on the equal number of incremental increases between one power of ten and the subsequent power of ten.

For example, the single digit incremental pattern between 0 and 10 (0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10) is mirrored by the pattern between 10 and 20, 100 and 200, 10,000 and 20,000, 100,000 and 200,000 and so on.

In addition, all the incremental series display a common pattern of internal symmetry. As a representative example, between 1 and 10 there are as many numbers below 5 as there are above. This aspect is crucial to rounding. To round a number to the nearest ten, the last digit of the number is the determinant. If that digit is 1 through 4, then the number is rounded to the next lower number that ends in 0 (the next lower even ten). For example, the number 14 is rounded to 10. If the last digit is 5 or more, then the number is rounded to the next higher even ten. For example, 17 would be rounded to 20.

The same pattern is followed when rounding numbers to the nearest hundred. Numbers that end in 1 through 49 are rounded to the next lower number ending in 00. Thus, 624 is rounded to 600, as is 648. Numbers that end in 50 or higher are rounded to the next even hundred. Thus, 650, 675 and 688 are all rounded to 700.

Similarly, when rounding to the nearest thousand, numbers whose last three digits are 001 through 499 would be rounded to the next lower number ending in 000, and numbers whose last three digits are 500 and higher would be rounded to the next even thousand. As examples, 4,390 and 4,450 are rounded to 4,000, while 4,600 and 4,835 are rounded to 5,000.

The same pattern carries through the increasing powers of ten.

DECIMALS

Rounding decimals is no more complicated than rounding whole numbers. If the thousandths place of a decimal (i.e., 0.00x) is four or less, that place is dropped and the value of the other digits remains unchanged. As an example, rounding 0.574 to the nearest hundredth produces 0.57.

If, however, the thousandths place is five through nine, then the hundredths place is increased by one digit. For example, 0.577 is rounded to 0.58.

Decimals can also be rounded to the tenth. In this operation, if the hundredths and thousandths places are less than forty-nine, they are deleted and the value of the tenths place remains unchanged. As an example, 0.638 is rounded to 0.6.

But, if the hundredths and thousandths places exceeded fifty, then these would be dropped and the value of the tenths place is increased by one. For example, 0.679 is rounded to 0.7.

Decimals that extend to more decimal places can also be rounded. Then, they utilize the preceding two decimal places to make the rounding decision. Thus, 0.647756 is rounded to 0.6478 and 0.32434612 is rounded to 0.324346.

PI

One prominent case is the value of the ratio of the circumference of a circle to the diameter of the circle. This ratio, which is called "pi," has been computed to approximately 100 billion digits, with no repeat or end of the decimal sequence found. The first few digits of pi are 3.14159265 . . . Since there is an infinite number of digits in pi, it has to be rounded to be used at all.

Using the aforementioned rules, pi could be rounded off to 3.142 (since the next digit is 5), 3.1416 (since the next digit is 9), or 3.141593 (since the next digit is 6). Obtaining all the rounded versions of pi would literally be a never-ending task. The decision of where to round off pi depends on the how precise the number needs to be, with more decimal places producing a more precisely accurate approximation of pi.

A Brief History of Discovery and Development

References to rounded-off values of pi can be found in the Bible and other sacred scripture. Many books contain apparently incorrect derivations resulting from the use of a rounded estimate of pi, indicate that rounding has been practiced since antiquity.

Real-life Applications

LENGTH AND WEIGHT

Rounding is a way of simplifying numbers to make them easier to comprehend and use. Precise measurements (at least to the nearest 0.1 inch) are necessary in carpentry to provide a correct fit between components. However, in many other applications an estimation of dimensions is sufficient.

Length measurements provide numerous examples of rounding. For example, a brochure advertising a house for sale might want to note that the house is far back from a busy highway. Instead of noting that the driveway is 221.5 feet long, rounding off the distance to 200 feet will still convey the desired impression to the prospective buyer.

Rounding off lengths can also be a more descriptive way of comparing two objects. As an example, it is accurate to describe a 39-inch-long board as being 1.77 times the length of a 22-inch-long board. However, this ratio is hard to conceptualize. Rounding off the lengths of the longer board to 40 inches and the shorter board to 20 inches produces a ratio of 2. This is close to the actual ratio, but is a much easier difference to understand.

This ease of comparison also applies to weights. An object that weighs 262 pounds is 4.3 times as heavy as an object that weighs 61 pounds. While descriptive, and certainly accurate, this weight distinction is more difficult to gauge than if the weights are rounded off to 250 and 50 pounds, producing a ratio of 4.

BULK PURCHASES

When contemplating the bulk purchase of an item, rounding off permits a quick estimation of the total purchase price. This can be an important factor in making the purchase decision.

An item may be advertised with a price (i.e., $10.65) based on a single purchase. In considering a bulk purchase (i.e., 21 items), rounding can be used in several ways. The item cost can be rounded (i.e., $11.00) and multiplied by the number of items to give the total purchase cost.

The number of items can also be rounded (i.e., 20). The purchase price can then be determined by multiplying the rounded single purchase price by the rounded number of items.

Either of the estimates, which typically are close to the actual (non-rounded) value, provides the information necessary for the purchase decision.

POPULATION

In a census, the population is determined as accurately as possible. However, such an exact tally is not always necessary or convenient. Indeed, in the case of a city of town, maintaining an exact tally can be difficult with the population shifting daily.

Rounding the population can be a more convenient way to present the information. This is especially true when the population figure can be rounded up, as would be the case for a civic population of 47,724. The rounded population of 48,000 would be a hedge for the natural increase in population number over time.

Many communities that post welcoming signage will use a rounded population number. At the very least, this saves constant modification or replacement of the sign to keep the population number current.

The rounding of population is also extends to the state level. For example, according to 2003 figures from the United States Census Bureau, the population of California was 38,484,453. Because this number will be constantly fluctuating with births, deaths, immigration and emigration, such an exact number may not be useful in some instances. Instead, the population can be rounded to the nearest million (38,000,000) or nearest hundred thousand (38,500,000).

Similarly, rounding can be done for selected categories of census data. Using the California example, the 2003 Census Bureau figures established that women made up 50.2% of the state's populations. Rounding the number to 50% makes discussion of the female segment of the population easier, without comprising the possible significance of the figure.

LUNAR CYCLES

"Thirty days has September, April, June and November," begins a well-known nursery rhyme. Calculations of lunar cycles are based on 30 days a month. This is despite the fact that there are only four such months in the year.

The derivation of the monthly period in our Julian calendar (which consists of three 365-day years followed by a 366-day leap year) is based on the monthly transit of the Moon around Earth. However, the lunar cycle actually covers some 29.54 days. The 30 day period that is enshrined in our calendar is a rounded estimate of the actual lunar cycle time.

Similarly, the 365-day length of the normal year—based on the transit of Earth about the Sun—is itself a rounded number. Because every forth year is a leap year to maintain the synchronicity of the calendar, each year actually comprises 365.33 days. Because the decimals are less than 0.50, the rounded number becomes 365.

ENERGY CONSUMPTION

Figures on the consumption of oil and gasoline that are released by agencies such as the United States Department of Energy (DOE) are rounded. For example, in 2004 the DOE reported that the United States used an average of 20 million barrels of oil per day. This number has been rounded up or down from a more precise estimate. The result is still a potent number, which serves as a reminder of just how much of a non-renewable resource is claimed.

WEIGHT DETERMINATION

Trucks that haul produce, freight and other loads on the nation's highways are designed to hold a maximum weight. Exceeding that weight can make a load dangerously unstable, which can lead to an accident. As well, trucks that exceed the weight limit for bridges and roadways can damage these routes.

To try to ensure compliance with weight limits, transport trucks are periodically required to pull into roadside inspection stations where the vehicles are weighed.

Part of the process used to determine weight compliance involves the establishment of several weight categories; for example, under 30,000 pounds, 30,000–80,000 pounds and more than 80,000 pounds. When a particular truck is weighed, the result can be rounded off to the nearest 10,000 pounds. So, a truck weighing 57,650 pounds can be recorded as 60,000 pounds. If the truck is meant to be in the intermediate weight category, then it is in compliance.

ACCOUNTING

When people successfully sell an item on the eBay Internet auction site, a portion of the sum they receive represents the company's fee. As more transactions are conducted, a running balance is kept of the amount owing.

eBay uses six decimal places to charge on an account. For example, a calculated balance might be \$6.333560. However, the balance tally that appears on a customer's account is rounded to two decimal places (\$6.33). As with other examples of rounding, there can be a slight difference between the actual and rounded values. But, rounding produces a balance that conforms to accepted billing practices.

Cash registers are programmed to round off a sum to the nearest hundredth. For example, if the sales tax charged on purchased items is 0.015%, then the sales tax added to an item sold for \$17.50 is 17.50×0.015, or 0.2625. The final purchase price for the item would be $\$17.50 + \$0.2625 = \$17.7625$.

Key Terms

Decimal: Based on the number ten; proceeding by tens.

Pi: The ratio of the circumference of a circle to the diameter: $\pi = C/d$ where C is the circumference and d is the diameter. This fact was known to the ancient Egyptians who used for π the number 22/7 (3.14159) which is accurate enough for most applications.

Whole number: Any positive number, including zero, with no fraction or decimal.

This four decimal place tally would never appear on the cash register receipt. Instead, the internal programming of the register rounds $17.7625 to $17.76.

TIME

Clocks and watches allow time to be measured to the second. With digital display capability, the estimation of time is needless.

Yet often time is rounded and expressed in more general terms. Instead of expressing a time as 6:43, a common practice is to note the time as "quarter to seven." Because an hour can be conveniently divided into four 15-minute segments, time can be rounded within a 15-minute bracket of time. Thus, a time of 6:32 could fairly accurately be rounded and expressed as "half past six."

The slight loss in the precision of the expressed time has not come at the loss of meaning.

MILEAGE

Mileage is typically expressed as the average of the distance traveled and the time taken to travel that distance. The result can be a multi-decimal number that is unwieldy and conveys too much information than is necessary.

As an example, if someone drove 335 miles in 4.75 hours, their average mileage would be 335 / 4.75, or 70.526315 miles/hour. For practical purposes, such as to calculate gas mileage, a simpler answer is best. Considering the above mileage to the first decimal place (70.5) allows the mileage to be rounded up to 71 miles/hour. The rounded answer carries the suitable depth of meaning for the problem.

Rounding is also practiced by drivers when calculating the distance from one local to another using a road map. Often maps will display distance to one decimal place (i.e., 25.5 miles). But, when adding a number of distances mentally, it is easier to round each of the distances to whole numbers and add the series of whole numbers together. This can usually be accomplished easily, quickly, and will provide the driver with the answer that has the appropriate level of meaning.

PRECISION

Precision is indicated by the number of significant figures in a number. For example, using a meter ruler, it is possible to measure a length to the millimeter (one thousandth of a meter, or 0.001). However, expressing some measurements to the thousandth of a meter can be imprecise.

For long distances, it would not be reliable or even honest to report a distance to this level of precision. Rather, by rounding the number to the tenth decimal place, a value is reported that represents a more reliable estimate of the distance.

Scientific notation is valuable in improving the precision of rounded numbers. For example, 363.6 meters can be expressed as 3.636×10^2. The number can also be rounded off and expressed in scientific notation as 4×10^2 or, more precisely, $3.6 \text{ x } 10^2$.

Where to Learn More

Books

Niederman, Derrick, and David Boyum. *What the Numbers Say: A Field Guide to Mastering Our Numerical World.* New York: Broadway, 2003.

Pickover, C.A. *The Mathematics of Oz: Mental Gymnastics from Beyond the Edge.* Cambridge: Cambridge University Press, 2002.

Pickover, C.A. *Wonders of Numbers: Adventures in Mathematics, Mind, and Meaning.* Oxford: Oxford University Press, 2002.

Web Sites

Arnold, Douglas N. "The Patriot Missile Failure" <http://www.ima.umn.edu/~arnold/disasters/patriot.html> (February 12, 2005).

British Broadcasting Corporation. "Skillswise factsheet: what is rounding?" <http://www.bbc.co.uk/skillswise/numbers/wholenumbers/whatarenumbers/rounding/> (February 12, 2005).

Purplemath. "Rounding and Significant Digits I." <http://www.purplemath.com/modules/rounding.htm> (February 12, 2005).

Rubric

Overview

The word rubric sounds like something from a poem. Indeed, the term comes from the Latin word *rubrica*, meaning 'red earth'. Historically, rubric refers to the prompts that were written in red ink in some documents during the Middle Ages. Later, red ink was used to highlight noteworthy sections (often rules) within legal documents.

But, contrary to the wordplay of poetry, a rubric is grounded in logic and math. A rubric is defined as a set of rules that allow tasks or activities to be scored. By doing so, both students and teachers are more aware of what achievement benchmarks there are in a task, and what a score really means in terms of what was achieved and what was not. This sort of knowledge can help students improve, since they are more aware of what specific tasks need to be done to improve.

Not surprisingly, the scoring involved in a rubric is tied to mathematics. The means used to assess the performance of the particular task and the detailed descriptions of the various levels of performance can involve math. Rubric math is certainly real life math, and is in action every day in most every classroom.

Real-life Applications

SCORING RUBRICS

A good rubric can detail out the various grades of quality of each of the criteria that have been picked as being important in the performance of a task. Often, this detail takes the form of a point scale. For example, in an oral presentation, a rating of 1 on a 1-5 point scale could be understood to mean 'a poorly organized presentation, with a poor use of voice and props.' A grade of 5 would be given for 'a presentation that is excellently organized and presented with a riveting use of voice and props.' The rubric has clearly detailed the expectations of the students and how their efforts will be scored. There is not any confusion over what a certain score means.

Another example will help to show the clarity that rubrics can provide. A poor rubric would be 'Students will show an understanding of how to use a ruler.' Instead, a proper rubric for this situation could read 'Students will demonstrate that they can use a ruler to accurately measure the target length and width of large and small objects.'

In the non-rubric classroom, scoring an assignment by a percentage value can lead to confusion. For example, two students hand in an essay on the same topic. One of

Captured former Iraqi leader Saddam Hussein undergoes medical examinations in Baghdad. Aspects of medical exams and other biometric tests can be converted to numerical data through the use of rubrics that allow examiners to match subjective observations such as gum color to a numerical scoring system. AP/WIDE WORLD PHOTOS. REPRODUCED BY PERMISSION.

the students receives a grade of 72 percent and the other student's grade is 76 percent. It can be difficult for a student to understand why one essay is four percent "better," and how another essay can be improved upon in the future, since percentage evaluations seldom have criteria associated with them. It may even be difficult for the teacher to explain the basis of the marking to the student!

Rubrics can also consist of questions. 'Does the student understand how to measure centimeters and millimeters?' and 'Can the student produce measurements that are two- and three-times as long as an example dimension?'

Scoring rubrics can also be used to gauge a student's performance. These can involve a checklist or using a number-based rating scale of performance. They specifically itemize the performance and provide a number that indicates how well or poorly a task was achieved. In other words, such rubrics provide quantitative results. Rubrics can also provide qualitative results. Examples include ratings like 'excellent,' 'good,' and 'poor.'

When different people look at an assignment or a test document, they can have different rankings of what are the important things to note. Scoring rubrics helps avoid ambiguity and differences in assessing an assignment or document, since the criteria for scoring will already be set out. Deciding on which criteria to include and in what order can be the toughest stage. But, once the rubric is established, scoring becomes routine.

There are several different types of scoring rubrics. Which one is used depends on the purpose of the evaluation that is to be done.

South Korean high school students bow in Seoul, as they pray for their seniors' success in college entrance exams. School exams that require essays, such as the revised College Scholastic Ability Test (SAT exams), are often scored by rubrics that define criteria that allow scorers to place numbers in otherwise subjective evaluations. AP/WIDE WORLD PHOTOS. REPRODUCED BY PERMISSION.

ANALYTIC RUBRICS AND HOLISTIC RUBRICS

These terms sound complicated, but they are not really. An analytic rubric is one that details how several different criteria are to be used in the evaluation of something. For example, an analytic rubric in an essay presentation could detail what aspects of grammar are going to be assessed and how the structure of the essay contributes to its impact. Each aspect could have a checklist of points that will be evaluated, and different scoring lists for the different criteria.

Sometimes it is not possible or beneficial to evaluate a project on separate criteria. Then, an overall view is best. That is where the holistic rubric comes in handy. In a holistic rubric a single scoring scale is used.

Generally, an analytic rubric is suitable for quantitative scoring (that is, where numbers are actually assigned as scores), whereas a holistic rubric is best suited for qualitative assessments.

The two rubrics are not mutually exclusive. It can happen that one of the criteria in an analytic rubric is more general. So, the assessment of that one criterion can involve a holistic rubric.

GENERAL RUBRICS AND TASK-SPECIFIC RUBRICS

In setting up the scoring of a task or activity, the first step is in creating what is called the general rubric. This is an overview or an outline that helps create the more detailed rules of the scoring. For example, if a course in school is designed to improve a student's skill in performing microbiology experiments, a general scoring rubric could be developed to evaluate each experiment that a student does. The feedback that a student got following the completion of one experiment could help them to carry out better experiments in the future.

A task-specific rubric would be concerned with the evaluation of an individual experiment. The criteria would be different from the general rubric, and would focus on that particular experiment.

A task-specific rubric lays out the details of how a single task is to be approached. As well, the rubric provides

a basis for scoring how well (or not so well) the particular task was done. Put another way, the task-specific rubric details what really counts about the particular task being done.

As another example, a rubric for an oral presentation could tell students that their presentation will be judged on the originality of the topic, the organization of the information, and the presentation itself; how informative and entertaining the delivery is; the use of voice; and the constructive use of props.

Developing a Scoring Rubric The very first step in developing a scoring rubric is figuring out what aspects of the project, report, lesson, or other item being evaluated are important to the evaluation. It is fruitless to focus on something that will not provide any feedback that can be used for future improvement.

The qualities that are identified as being important form the framework on which the rubric is made. For example, in assessing a report such qualities could be spelling, grammar, organization, presentation style and use of language. The details of the rubric would be compiled using these as the starting points.

There should be enough qualities to make for a meaningful assessment, but not too many qualities. If there are many qualities, it can become difficult to score any one of them. It is better to have fewer qualities with several scoring criteria in each one than a lot of qualities with only one criterion in each.

Ideally, there should be three criteria per criterion, since typically there will be indicators of poor, average, and standout performance for each quality. The criteria should not depend on each other. Each should be able to be evaluated on its own.

When developing the criteria, it is better to have a definite indication of how each criterion will be determined. For example, it is better to say 'Student's writing will be free of spelling errors,' than to say 'Student's writing will be good.' 'Free of spelling errors' is something that can be quantified. 'Good' is hard to quantify.

If the evaluation involves assigning a score (1,2,3, . . . or A,B,C, . . .) then the same score should mean the same thing for different categories. It would be confusing to have a score of 2 pertain to merely satisfactory in one category and outstanding in another category.

Finally, the rubric needs to be tested in action. Typically, the first run-through of a rubric will show that some revision is necessary. This is to be expected. The math involved in a rubric is not the more straightforward math of an equation. Rather, the mathematics of scoring is part of a more subjective evaluation. So, some tinkering may be needed to make the rubric as good an instrument of assessing performance as it can be. But, the effort will be worth it.

The math that is part of a rubric can help create a tool that assesses the performance of a task in a way that is clear to the teacher and, most importantly, to the student. The student will be able to use the information to improve. Different teachers will be able to use the same rubric effectively. The real life math of a rubric is thus an important part of a great classroom.

Where to Learn More

Books

Arter, J.A., and J. McTighe. *Scoring Rubrics in the Classroom: Using Performance Criteria for Assessing and Improving Student Performance*. New York: Corwin Press, 2000.

Moen, C.B. *25 Fun and Fabulous Literature Response Activities and Rubrics: Quick, Engaging Activities and Reproducible Rubrics the Help Kids Understand Literature*. New York: Scholastic Professional Books, 2002.

Web sites

Middleweb.com. "Just what is a rubric?" <http://www.middleweb.com/CSLB2rubric.html> (November 9, 2004).

Moskal, B.M. "Scoring Rubrics: What, When and How?" <http://pareonlne.net/getvn.asp?v=7&n;=3> (November 9, 2004).

Smith, J. "Base Arithmetic." <http://www.jegsworks.com/Lessons/reference/basearith.htm> (October 30, 2004).

Sampling

Overview

Sampling is the statistical process of analyzing a group of items selected from a bigger set. It is not always feasible to perform a study on all members of a population. For example, it is impossible to interview all AIDS patients throughout the world to study the stress and problems they endure in their daily life. A better approach is to interview a group of such patients and generalize the common results to all HIV positive people, the world over.

In sampling, a smaller manageable group of items, elements, members, or individuals representing the entire population is studied. Observations made from the analysis are generalized for the larger set to which the sample belongs. Sampling helps identify and understand the group's dynamics, ongoing trends, and their implications.

Sampling is a widely implemented concept used to perform demographic studies, environmental research, marketing analysis, and soil testing, to name a few applications. It will not be incorrect to say that sampling is utilized in all aspects of life ranging from medicine, social behavior, business, music, sports, and technology to ecology, and the balance in nature.

Fundamental Mathematical Concepts and Terms

PROBABILITY SAMPLING

Probability and non-probability sampling are the two commonly used forms of sampling implemented in various sciences. In probability sampling, every member (or object) of the sample group gets an equal opportunity (in other words, they are all given equal weight). Probability sampling begins by listing the traits and features to be studied. Identifying these traits helps in defining the populations to be researched. For example, to study the effects of smoking on women of reproductive age, female smokers in the age group of twelve to fifty years are most likely to show traits identifying the sample to be studied. If however, this study is to be conducted for women from a particular ethnicity or geographic region, then the subject's background and location will also form a part of the features defining the population to be analyzed.

Once the group to be studied is identified, all individuals belonging to that group have the same opportunity to participate in the research effort, thus reducing bias and error. However, at times, scientists randomly choose their subjects from the selected group. This constitutes unrestricted or simple random sampling.

Another type of random sampling is restricted or stratified sampling, in which the population is categorized into homogeneous segments with the idea that maximum possible variations can be accounted for, thereby minimizing the chances of arriving at biased results. Samples representing each unit are then identified and studied.

In contrast to random sampling is the more frequently used systematic sampling, wherein the first element is selected randomly and the remaining elements are identified on the basis of a calculated sampling interval.

For example, a student might want to interview store owners of all the malls in a particular location. If the identified area has several malls, then it would be extremely time-consuming to talk to all the owners of all the stores of all the malls. To make the task easy and still get a good representation of the owners, the student can determine the total number of malls and stores. Assume there are a total of ten malls and 250 stores. The student decides to interview 20% of the population, which means fifty vendors. The sampling interval can be easily computed by dividing population size (250 in this example) by the sample size (50 in this example). Accordingly, the sampling interval in the illustrated example is five.

Another form of probability sampling is cluster sampling, in which the investigator selects subjects in a "phased manner"; first identifying clusters to be studied, and then randomly or systematically identifying individuals to participate in the study. Using the example discussed earlier, a researcher may, for example, first identify cities from where to select women smokers to perform the study. The examiner then randomly or systematically selects individual participants from the identified locations.

In the real world, none of the sampling techniques mentioned above are employed in isolation. Instead, researchers use a suitable combination to perform studies. This strategy of utilizing one or more techniques to investigate an issue is known as multi-stage sampling. It is helpful in carrying out elaborate research involving huge populations spread over large geographical areas.

NON-PROBABILITY SAMPLING

In non-probability sampling, the researchers typically select subjects depending on their availability. The basic assumption of non-probability sampling is that any sample available would be sufficient to accurately represent the entire population, thus leading to correct results. With non-probability sampling, not all members of the group receive an equal opportunity to participate in the study.

Some forms of non-probability studies are conducted with individuals easily available. For example, visitors to malls or other public places might find a television crew interviewing passers by, thus offering an inexpensive way of understanding public opinion on a particular subject. This form of non-probability sampling, ideal for quick, economical, investigative, and narrative researches, is therefore referred to as convenience, accidental, or haphazard sampling.

The biggest disadvantage of convenience sampling is the high degree of bias because it is completely at the analyst's judgment to select members for the study. However, errors occurring due to bias can be minimized if the study is conducted on a uniform population that shows consistent features through its expanse. This way, all members experience and perceive almost the same things and represent an accurate picture. For example, soil samples from a smaller field are likely to yield similar results, but the probability of samples collected from a large field showing greater variation is high.

Another form of non-probability sampling is volunteer sampling that subjects a volunteer to participate in the study. A good example of volunteer sampling is the call-in surveys that various television and radio channels conduct to assess public sentiment. The downside of volunteer sampling is that only those interested enough in the issue participate in the study, thereby introducing strong biases. It does not take into account the opinions of those who might be concerned about the issue but avoid active participation.

Yet another method of non-probability sampling is judgment sampling, wherein the investigator uses his/her judgment to select the members for a study. The biggest disadvantage of this sampling method is that the selector's judgment might be heavily biased and inaccurate. In such a situation, results of the investigation carried out will be erroneous, no matter how elaborately the study is performed. These biases can however be reduced if judgment sampling is utilized in controlled environments, such as a life sciences laboratory where variables are few and limited by the issue to be studied.

Apart from these types of non-probability sampling, the most commonly used is quota sampling, in which members are selected from various sub-groups of a population until they satisfy a pre-calculated quota that is typically in proportion to the total population size. Several market researchers use the principle of quota sampling. For example, a mattress manufacturing organization might want to know the opinion of senior citizens who constitute 5% of the population. If the sample to be studied is of 1,000 people, then of those 1,000 candidates, 50 must be senior citizens. To meet this quota, the interviewer will then approach any 50 people fulfilling the age criteria.

Snowball sampling, another form of non-probability sampling, involves employing few subjects for a study. These members, in turn, enlist their acquaintances (people they know), who on their part sign up their friends and colleagues for the study. The basic idea here is that the individuals who have signed up initially would be the best to know about more of their kind. For example, a quilt supplies store aiming at updating their offerings in accordance with the new products available in the market, and the precise needs of their patrons, will be better off by interviewing quilters. Here, the best way to interview the maximum number of people would be to request regular customers to suggest names of fellow quilters they might know.

As is obvious, snowball sampling is useful in studies targeting small or inaccessible populations. The situation may further become difficult if the members of such populations are scattered everywhere.

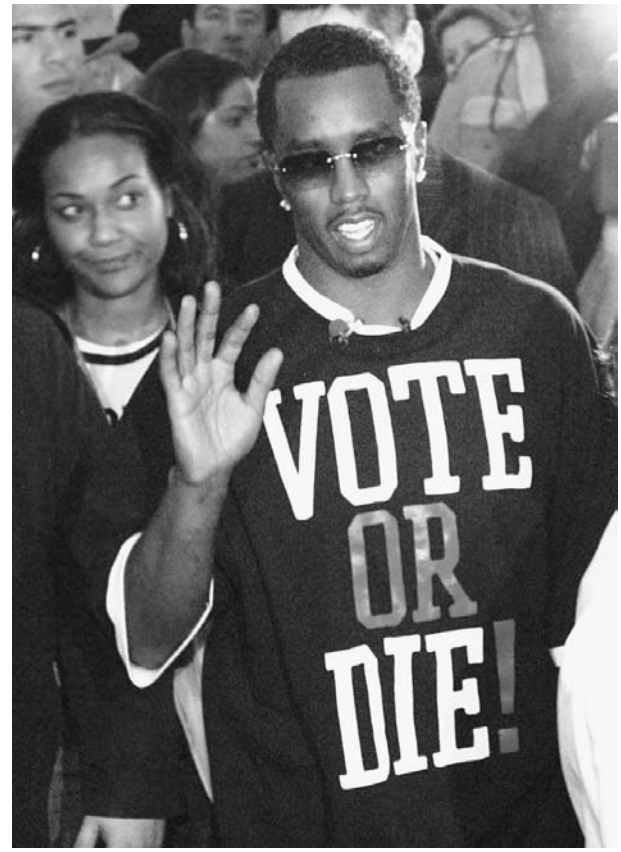
A Brief History of Discovery and Development

Sampling is almost always a part of scientific testing. Rarely are all objects or situations examined, so by testing selective samples, scientists are able to make broad conclusions.

It was only through careful sampling and segregation of pea plants that Gregory Mendel (1822 A.D.–1884 A.D.), known today as one of the founders of modern genetics, discovered the laws of heredity. His methods since then have been followed to produce improved varieties of not only crops and plants but also award-winning breeds of dogs, horses, cattle, and other animals.

Similarly, Englist naturalist Charles Darwin (1809 A.D.–1882 A.D.) studied samples of animals and birds living on the Galapagos Islands. Diligent analysis of those limited population samples resulted in the theory of evolution, the unifying principle of biological science.

Starting from 1920, the *Literary Digest*, in circulation from 1890 to the late 1930s, correctly predicted the winner of the presidential campaigns for four elections in a row. Their method was simple. They collected the name of the voters' favorite candidate in six states. They however subsequently failed the fifth time because they selected samples from telephone directories and auto registration records, thus approaching only the wealthy strata of population. Beginning 1936, Gallup presidential polls used quota sampling to successfully predict presidential elections. Today, politicians and pollsters use sophisticated mathematical sampling to predict elections and to shape policy.



Sean "Diddy" Combs and others used sampling to determine where best to target efforts to encourage young people to vote during the 2004 presidential election. AP/WIDE WORLD PHOTOS. REPRODUCED BY PERMISSION.

Real-life Applications

AGRICULTURE

Sampling has been implemented to improve agricultural practices since ancient times. It is a well-known fact that not all kinds of plants grow in similar kinds of soils and climates. Meticulous sampling and testing over the centuries has now made it clear the type of soil a crop requires for thriving.

Today, samples are collected and put through extensive tests to assess a variety of factors affecting plant growth, soil quality, pH balance, nutrient levels, and concentration of micro-organisms. As mentioned earlier, most real-life applications use a combination of sampling methods. The same is the case here. However, the most prominent sampling method used in agriculture is the cluster sampling, wherein groups are identified and then they are either systematically or randomly used as samples.

PLANT ANALYSIS

Sampling analysis of plants is useful in identifying any nutrient deficiencies or excessive accumulation of a particular nutrient that proves harmful to plant growth. By collecting samples of plant tissue and studying them carefully, scientists can also assess the effects of insecticides, pesticides, new chemicals thought to have beneficial effects, and the proximity of an industrial area or waste dumping grounds on the general health of the plants of a specific region.

Plant analysis can give crucial insight into the kind of plants a given piece of land will support in the future. For example, a random plant analysis of a green area near a waste chemicals removal yard will clearly show that the existence of the greenery is threatened. From this, one can easily deduce that the land might become completely barren in a few years and may actually become a wasteland. Accordingly, the authorities can take measures to save the fertile land.

Similarly, random sampling of plants showing a sudden drop in development can give important clues about the cause of the weakening. The reasons can be numerous, ranging from introduction of a new insect or bug into the plant community and increased human interference to a sudden rise in the local temperature. All this investigation however depends heavily on a best possible collection of enough samples representing plants of the identified region.

Sampling enough plants gives a good indication of the general health of other similar plants in a field. The entire sampling process followed here is based on the mathematical concept of cluster sampling. Clusters are first identified. For example, plants near a waste chemicals removal yard form one cluster; plants showing a sudden drop in development become another cluster, and so on. Random samples from each of these clusters are then studied.

SOIL SAMPLING

A regular soil analysis helps farmers and people engaged in commercial agriculture in assessing the quality of the soil. Depending on the findings of the soil analysis, they can enrich their fields before sowing the next crop. Alternatively, if the analysis shows increased occurrence of a harmful element, they can take due measures to prevent its growth.

Plants continuously absorb minerals and other nutrients from the soil. An annual soil analysis before the next sowing can tremendously help farmers in selecting the most appropriate crop to be grown and the right fertilizer to be used, thereby bringing down the costs, reducing avoidable harmful effects of unsuitable chemicals, and increasing crop yield to the maximum. All this leads to enhanced profits in terms of resources as well as money.

Typically, a soil analysis requires systematic sampling requiring the collector to visualize the field to be divided into grids. Based on a sampling interval is determined. For example, samples are gathered at a rate of 15–20 soil samples from a 40 acre field. Collecting fewer samples runs the risk of producing inaccurate results. To investigate nutrient concentration, especially that of nitrogen, samples from varying depths are taken. In other words, the sampling interval is critical.

SCIENTIFIC RESEARCH

Modern research depends largely on intelligent ways of implementing sampling procedures; so much so that no exploration can be carried out without collecting and analyzing samples. Numerous ongoing studies are a case in point.

By carefully studying samples of available material, scientists have discovered microorganisms thriving in the most extremes of locations, such as hot sulphur springs that have very low levels of oxygen. These studies involve a combination of cluster sampling with systematic and stratified sampling. Interestingly, the evidence supporting this discovery is so strong that the theory about all life needing oxygen for its sustenance is now being questioned.

Clinical trials testing effects of new drugs on human patients are possible only because of large numbers of patients volunteering for the tests. A combination of volunteer and judgment sampling are used to conduct these tests. Volunteers are required for the tests, and selection of the volunteers depends on the judgment of the person doing the sampling.

Geologists also depend on regular random sampling to continue with their explorations of Earth.

Scientists from all disciplines of studies put in collective efforts to discover any evidence throwing light on living beings that walked the surface of Earth before humans. Fossil samples are studied to find missing links in the evolutionary process. This process is based on concepts of stratified sampling. Dinosaur fossil studies can prove to be the key in determining the cause of their sudden extinction.

DRUG MANUFACTURING

Sampling is an extensive and essential part of the process of designing and manufacturing drugs. Right from the beginning, research assistants accompany surgeons in operation theaters to collect samples of diseased

tissues and conduct experiments to determine the cause of the ailment. As a part of these experiments, they compare the symptoms of a diseased animal with those of a healthy animal. To arrive at an authentic conclusion, they use judgment sampling to maintain batches of ailing and healthy tissue so findings can be generalized for the human population. However, this application requires both cluster sampling to identify the samples, and then analyzing the selected samples based on judgment.

Supporting these studies are the patient interviews or surveys regularly conducted by people involved in health care management and social sciences. Such investigations give the emotional, societal, and mental perspective to the effects of the disease. Depending on these studies, communities may shun or accept patients of a particular disease. For example, the stigma with AIDS is so strong that even the family members of an HIV+ person begin to avoid him or her, though everyone knows that AIDS does not spread through merely touching, hugging, or talking with the patient. This misconception however is gradually losing ground after countless interviews, medical reports, and proper promotion of facts about the basic respect that HIV positive people crave.

After understanding the causes and symptoms of a disease, scientists propose solutions and conduct experiments to study the practicality of the proposed remedies. This time round, carefully selected samples of diseased animals are treated with chemicals or therapy thought to be beneficial and the results are studied. Different treatments may affect different aspects of the disease. For example, a prospective drug to combat hair loss may influence hair growth on the whole body while another may have localized effects on certain body parts.

In addition, it is also important to study the side effects of these treatments. For instance, the so-called hair growth promoting medicine with localized effects may sound good, but might cause extreme nausea and dizziness, in which case it is not user-friendly and loses its marketability. On the other hand, researchers might accidentally stumble upon some positive side effects. A probable flu medicine may help reduce obesity, in the event of which drug manufacturers can either start a new investigation exploring the weight-reducing effects of the drug or market the flu tablet as it is while highlighting its desirable side effects too. Each of these can be thought of as clusters that undergo specific types of mathematical sampling.

Once the drug is ready for human trials, patients volunteering through non-probability sampling are invited to try it out, and studies are performed vigilantly to assess its impact. Volunteer sampling is the key here because members chosen through other sampling methods would literally force them to try the medication and thus prove to be unethical.

Quite often, new medication may not show immediate side effects, but its long-term use may cause unwanted results. Therefore, extensive data spanning several years of drug consumption are maintained and further examined. If possible, patients are grouped by different criteria, including age, sex, race, and region, to identify if a particular population responds uniquely to the treatment.

In a nutshell, independent of the stage of drug manufacturing, sampling is an important step in the introduction of new cures. Different forms of sampling may be employed at different times but it would be impossible to pioneer new drugs without the painstaking task of assembling samples of tissues and subjects that match the required factors—a concept based on the statistical principle of cluster sampling.

WEATHER FORECASTS

Weather predicting organizations receive loads of related data, which is then used for weather forecast. In spite of all the predictions carefully arrived at, the actual weather conditions are invariably somewhat different. Weather information analysts now compare samples of this difference along with current information to statistically arrive at the best weather forecast model for the next day, week, or month.

In the event of an approaching snowstorm or a thunderstorm, its estimated force can be compared with a past event of similar nature. Information from such a study is used to caution the public about an oncoming natural catastrophe, and estimate the extent of loss. This can be thought of as a type of convenience sampling, where information is presented based on availability (in this case, past data in a similar situation).

A different form of sampling referred to as matched sampling is used to calculate the risk ratio of accidents occurring on a bad weather day, particularly those related with some form of precipitation. In this type of sampling, a given period of unfavorable weather conditions is matched with the same duration of otherwise desirable weather. For instance, a heavy snowfall period starting at 9:00 a.m. on a Friday morning lasting two hours is compared with an identical two-hour period of another Friday morning with clear weather conditions. The control duration is ideally selected from a couple of weeks before or after the time period to be studied. Number and types of accidents occurring during the experimental duration are compared with those occurring during the control period to arrive at risk ratios. Comparing a specific weather related duration with a control is essential in matched sampling.

ENVIRONMENTAL STUDIES

Sampling finds widespread use in environmental studies, especially those related with measuring pollution. To study air and water pollution levels at different places, researchers collect samples and put them through numerous scientific procedures to draw conclusions.

Levels of pollutants in air and water can be studied from various perspectives. For example, assessment of harmful airborne microbes is particularly useful for food processing units, organizations handling any kind of living organisms, pharmaceutical companies, and hospitals. Larger air samples are required from places considered having relatively cleaner air, because they have fewer pollutants.

Studies indicate that collecting air samples and testing them is a far better and more accurate approach than traditional methods. Additionally, sampling is quicker and can be done in a shorter duration.

Similarly, studying water samples helps identify water contaminants following which corrective measures can be taken to save the water body. Trained personnel and investigators take samples of water from the source to be examined. Random water samples, collected in special containers, are carefully scrutinized to determine healthy and harmful elements contributing to the general health of the water resource.

It is interesting to note that results of water sampling may differ depending on the time when samples were collected. For results to be an accurate representation, due care must therefore be taken to collect all samples at the same time.

Environmentalists and ecologists regularly assess the delicate balance maintained within a system by performing various tests on the objects and living beings specific to that natural environment. For example, researchers evaluate soil quality of an exceptionally fertile area by studying samples of its soil, water, flora, and fauna. Similarly, productivity of fishing grounds can be judged by doing water analysis and a study of the fish inhabiting the place.

All of the above are processes based on sampling. Some involve random sampling, while some are applications of stratified sampling. The foundation of each of these is, however, identifying different groups of samples—similar to what is done in cluster sampling. Thus sampling in environmental studies can be thought of as a combination of cluster sampling with other kinds of probability sampling.

DEMOGRAPHIC SURVEYS

Demographic surveys involve counting the number of people who match the criteria to be studied. The most well known form of demographic survey is the census, wherein the government launches a mass scale activity, typically every ten years, of counting the number of people living in the country.

Though sampling is not an inherent part of census, it was used in 2000 to calculate the number of people belonging to minority groups, and the homeless since they lack an address. This involved a mix of judgment sampling, quota sampling, and convenience sampling.

Once the total population is calculated, different types of sampling, be it probability or non-probability sampling, are used for various purposes. For example, comparison of samples from the latest and past census results can be used to assess ongoing trends within the country. A study in the early 2000s showed that an increasing number of women drivers are getting involved in road accidents. Though this study may throw some light on the increasing stress levels among women drivers, but one must also note that over the years, more and more women are acquiring driving licenses. In fact, the number of women drivers is increasing faster than ever before. This, to a large extent, explains more women being involved in accidents.

Census results also give key information about future requirements and society make up. For instance, if a given region shows increased levels of education, it can be safely assumed that people from that region have a higher probability of performing well in the future. If on the other hand, census results from a flourishing area show an increase in the number of elderly people and a relatively lower number of young adults and children, then that area may suffer setbacks or show diminished development after a few years.

Demographic surveys can thus be used to deduce population dynamics. The objectives of such analyses range from identifying problems of minority groups, advancements in society, general progression of a group of people, and population health to predicting growth rate in the coming years and expected development responsiveness.

ASTRONOMY

Astronauts have brought back samples of soil and rocks from the moon. Because of the bulky spacesuits the astronauts wear, their mobility is quite restricted. To enable them to successfully collect samples, NASA has used a variety of tools such as improvised rakes, tongs, scoops, hammers, and electric drills on their trips to the moon.

The procedures that are to be followed to collect these samples depend on several factors. For example,

while collecting samples from near a crater, astronauts follow what is known as radial sampling (a mathematical concept). The basis of radial sampling is that materials thrown out of greater depths are deposited closer to the rim of the crater, while substances coming out of shallower depths accumulate far away from the rim. The astronauts therefore collect samples from varying distances from a crater, thereby ensuring collection of matter from different depths from the surface.

Similar to the astronauts, space vehicles and rovers have been involved in collecting rock samples form the surface of Mars. While the space machines collect and send the samples back to Earth, scientists in laboratories study them to derive conclusions about Mars.

Mars samples include photographs assisting investigators in identifying objects of interest. High-quality electronic imaging equipments are required for this purpose. Later, after identifying the rocks to be examined further, sophisticated drills and other instruments are used to retrieve the appropriate samples. Somewhat varying from astronauts and rovers, but essentially achieving the same objective, are space shuttles that continue to send pictures of planets and other objects in the universe. A thorough examination of these pictures and their comparison with collected data unfolds new information literally everyday.

Aside from this, astronomers use standard sampling procedures to devise theories about the universe. Sampling plays a key role in the development of astronomy as a science because it is impossible to perform laboratory experiments the way it is possible for other science streams. Astronomers must use sampling to draw inferences about changes going on in the universe.

Sampling further becomes feasible in astronomy because any event in the universe takes millions of years. However, different objects in the universe can be observed in various stages of the event, thus making it possible for astronomers to predict the complete cycle.

ARCHEOLOGY

Archeologists use cluster sampling to identify and decide the sites to be dug for archeological findings. If an area of interest shows ancient artifacts on the surface, such as the arid and semi-arid regions, it is easy to explore and ascertain the particular spots to be excavated.

A problem, however, arises if objects of archeological value are embedded deep under Earth's surface and are covered with soil, grasslands, ice, human neighborhoods, and other objects on the ground. In such a situation, archeologists typically utilize the principle of probability sampling to pinpoint the specific areas to be dug up.

Systematic sampling is used to divide the region of interest in a grid-like fashion into excavating units that lie adjacent to each other without overlapping. Of these units, those identified to be explored further are dug with the help of machines or sometimes manually to detect any findings of an archeological nature. Measures are taken to employ non-destructive, though more time-consuming, digging methods.

MARKET ASSESSMENT

Marketing is the heart of any profit-making business and all businesses aim at making huge profits year after year. In spite of this, some companies perform exceptionally well while others fail to even capture consumer attention.

One of the key reasons for varying business performance is an understanding of the market. Organizations generating huge revenues typically have a sound understanding of their customers' needs. In other words, they have a grip over the market, possible only through market assessment.

There are different ways of studying market trends and all of them employ various sampling techniques. You might have come across company employees, typically students and other part-timers, working from a stall in a mall or a department store promoting a new product or collecting feedback from the visitors. People interested in the goods being endorsed or the parent company always make it a point to spend a few minutes at the stall either to know more about the product or give their input. This is an effective sampling strategy as it not only gives the customers the option of interacting with the organization, but the manufacturer also comes in direct contact with the target clientele. Gathering information otherwise can be an expensive, time-consuming activity that might drain an organization's resources.

The success of sampling depends largely on careful planning and interpretation of the collected data. Before embarking on such an activity, analysts should first identify objectives of the study, design a robust sampling process that includes defining characteristics of the population to be studied, the best suited method of selecting subjects, the ideal number of members constituting the sample, and the most effective way of conducting marketing research and assembling information. Additionally, researchers should devise a competent way of analyzing all the accumulated data. Without a serious approach to conducting market research, even the most extensive sampling strategy fails.

MARKETING

Several companies promote their new products by flooding the market with samples of the product.

Key Terms

Convenience sampling: Sampling done based on the easy of availability of the elements.

Simple random sampling: A sampling method that provides every element equal chance of being selected.

Stratified sampling: In this type of random sampling, elements are grouped together before sampling.

Systematic sampling: In this type of sampling, there are intervals between each selection for sampling.

This gives the consumers an excellent opportunity to try out the product in little quantities at nominal rates. If the users like the product, it is easier for them to switch to the regular packing. Sampling thus gives the producers a chance to gauge public response while users test the new product.

Even for products that have a strong market share, companies often retain their samples so retailers can offer them to prospective clients. Cosmetics manufacturers frequently use this marketing technique. Their outlets promptly offer sample sachets or trial packs to interested customers. Otherwise too, travelers go for smaller packing because of their convenient size.

For goods targeting a specific section of the market, producers send samples through mail. Organizations involved in producing baby products, for instance, send sample formula, diapers, diaper rash ointments, shampoo, and parenting magazines to new parents and parents-to-be. Similarly, crafters receive samples of new crafting products specific to their craft.

Typically, receivers of these samples constitute a special group of consumers for their needs form a niche market. They usually first join the manufacturers' mailing list or at least show an interest in trying out the products by filling out a form either on the Internet, or in a magazine, newspaper, or some other source. This way the organizations making specialized products get in touch with the clients who are actually interested in their products, without spending money in extensive advertising targeting everyone in general.

Offering samples is a win-win situation for both the consumers as well as the producers. While the clients can test the product either for free or at reduced prices, manufacturers reach out to consumers who are truly interested in their goods and are instrumental in giving them feedback.

Where to Learn More

Books

Thompson, Steven K. *Sampling.* Wiley-Interscience, 2002.

Thompson, William. *Sampling Rare or Elusive Species: Concepts, Designs, and Techniques for Estimating Population Parameters.* Island Press, 2004.

Web sites

NC State University. "Sampling" <http://www2.chass.ncsu.edu/garson/pa765/sampling.htm> (May 09, 2005).

Scale

Overview

Scale defines the relationship between the actual size of an object and its representation in the form of a prototype. It is used extensively in a variety of real world scenarios and is useful in modeling extremely large objects (or even tiny objects) into an easy to comprehend size. Scaling is done with respect to certain properties of the object, such as length, temperature, or mass. This concept is employed widely in architecture, astronomy, and imaging. For example, a scale model of a part of the solar system can provide a clearer understanding of the relative size and distances of planets and other objects in it.

Scale is also used in numerous other aspects of daily life including music, art, sports, fitness, business, technology, aviation, and a whole range of sciences, such as physics, chemistry, and engineering. The most common application of scales is found in maps.

Fundamental Mathematical Concepts and Terms

LINEAR SCALE

Scales can be associated with various properties of an object. Accordingly, there are several types of scale. The most basic form of scale is the linear scale. The linear scale follows a linear pattern and is used to quantify distance. The foot-ruler and the measuring tape are most well known examples of linear scale.

A key characteristic of the linear scale is that the length represented by two equidistant marks is always the same. Take, for example, a scale marked as 100, 300, 500, 700, and 900. As shown below, this would be a linear scale, as the length between any two equidistant marks, say 100 and 300, or 700 and 900, is always 200. (See Figure 1.)

Maps are the most prominent applications of linear scale.

LOGARITHMIC SCALE

One of the biggest limitations of linear scale is that it becomes difficult to manage if the quantity (or length) represented by the scale has a large range. This is where the

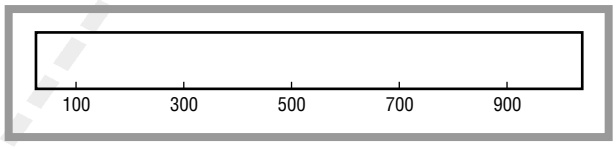


Figure 1: A linear scale.

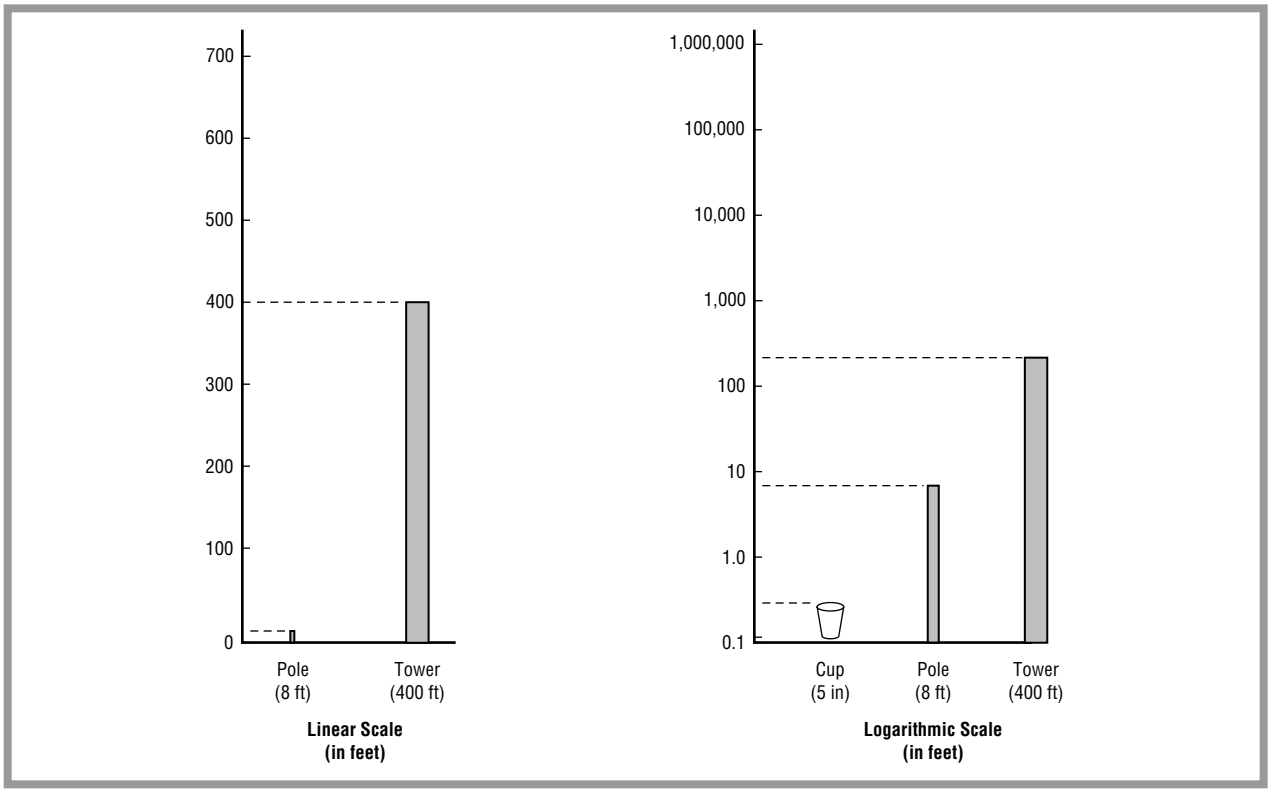


Figure 2.

logarithmic scale comes in. Simply put, the logarithmic scale represents the logarithm of the quantity rather than the quantity itself. In other words, the natural steps on a logarithmic scale increase in a multiplicative fashion rather than an additive or linear fashion. For example, a scale marked as 50, 500, 5000, 50000 would be a logarithmic scale as any succeeding mark is ten times the preceding mark ($50 \times 10 = 500$, $500 \times 10 = 5000$, and so on).

Differences between linear and logarithmic scales is shown in Figure 2.

On the linear scale in Figure 2 it is not possible to indicate the cup. The 8-foot pole is also not accurately shown. However, all three objects—the tower, the pole, and the cup—are clearly indicated on the logarithmic scale. Logarithmic scales thus become extremely useful in such scenarios.

INTERVAL SCALE

In an interval scale, the adjacent points are at equal intervals and also represent the magnitude of the underlying quantity. The most common example of interval scale is the thermometer (for measuring temperature in Celsius). The difference between 15°C (Celsius) and 20°C is the same as that between 30°C and 35°C—5°C. In other words, we can say that 20°C is warmer than 15°C by 5°C, and similarly for the other two points.

However, the interval scale does not have an absolute zero point. Consequently, we cannot say that 30°C is twice as warm as 15°C. The reason for this is that 0°C is not "absolute" zero (there is some heat at this point as well). The Celsius thermometer, thus, qualifies as an interval scale application.

In comparison, a scale indicating the percentile values of a few students may not qualify as an interval scale. The percentile value specifies the percent of total distribution that is equal to or less than that value. For example, in our case, saying that a student scores in the 75th percentile would indicate that 75% of the total students are either ranked equal to or below this student. Consequently, the difference, in terms of number of students between the 80th percentile and 75th percentile, may not be the same as between the 50th percentile and 45th percentile.

RATIO SCALE

The ratio scale includes all features of the interval scale. In addition, it also has an "absolute" zero or "true" zero point. Such scales allow for measurement of magnitude between equal intervals (as in the case of interval

scale) and permit calculation of ratio between two points as well. As a result, we may say that a certain value on the ratio scale is twice as much as another value.

Take, for example, the Kelvin temperature scale. Due to the presence of an absolute zero—0 K (Kelvin), 100 K is certainly twice as warm as 50 K. Moreover, the difference between 15 K and 20 K is also same as the difference between 40 K and 45 K.

The Kelvin temperature scale can also be used to show how our earlier statement (30°C is not twice as warm as 15°C) is true.

For example 0°C = 273 K, yet 30°C = 273 K + 30 = 303 K and 15°C = 273 K + 15 = 288 K. Accordingly, 303 K is not twice of 288 K and 30°C is not twice as warm as 15°C.

NOMINAL SCALE

The nominal scale is perhaps the most primitive model of measurement. It is a classification tool more than anything else. For decades now, mathematicians have been questioning the authenticity of this scale. Nevertheless, we will still discuss this concept in brief.

The numbers presented on a nominal scale do not indicate values or quantity. They simply indicate a category or a type. For example, consider a herd of different animals including bears, elephants, tigers, and lions. All bears may be categorized as "1" on the nominal scale, all elephants as "2", tigers as "3", and lions as "4". The intervals do not signify anything in terms of magnitude or quantity. We cannot say that elephants are twice something as compared to bears. In other words, the difference in various points on the scale cannot be measured in amount, but only in kind.

Consequently, none of the mathematical operations, such as addition, multiplication, subtraction, division, and average can be applied to this scale.

ORDINAL SCALE

The ordinal scale improves on the nominal scale as it adds more value to the different categories represented. Categories can be ranked or logically ordered on the scale based on certain characteristics. In a nutshell, with an ordinal scale you can say whether a particular item possesses more or less of a characteristic as compared to other items.

To make more sense out of this scale, higher values are usually represented by higher numbers. An example of ordinal scale is as follows. You visit Orlando, Miami, and Tampa along with your family and at the end of your stay are asked to rate each of the hotels (on a scale of 1 to 5) with respect to certain characteristics. These could include quality of the room, quality of service and staff, facilities provided, proximity to Disneyland, and so on. You rate the hotel in Orlando as "4", the one in Miami as "2", and Tampa as "2". This becomes an ordinal scale where you can conclude that the hotel in Orlando is better than those in Miami and Tampa.

An important point to note is that the intervals on the ordinal scale may not be equal. Similarly, you cannot establish a ratio between two values. For instance, in the above example, it may not necessarily mean that the hotel in Orlando is twice as good as the other hotels.

A Brief History of Discovery and Development

Applications of scale can be seen in some of the very ancient paintings and maps. Researchers have found evidence that maps based on certain scales were being used more than 2,600 years ago. In 1963, an interesting painting—dating back to 6200 B.C. was discovered in the city of Ankara, Turkey. The painting depicted a miniature version of a city known as Catal Hyuk (part of modern Turkey) in great detail. The painting included streets and houses of this town. This is one of the first known examples of scale. There have also been discoveries of other maps in the regions of Egypt and Greece between the periods 2300 B.C. to 600 B.C.

Around the same period, scale became an integral component of architecture in Greece, Egypt, and Rome. Through the centuries, there has been evidence of scale being used to design and build monumental structures. Musical instruments that date back to the early 1500 A.D. were also designed using scale theory.

In the late A.D. 1700s, the French Academy of Sciences devised a more consistent and organized unit of measurement. This new unit, known as the "meter," was based on multiples of ten—a concept commonly used in modern scales. Eventually, scale also found use in other fields, including model railroading. By the beginning of the nineteenth century, people started using scale commonly for a number of purposes with respect to business, sports, and a host of sciences.

Real-life Applications

MAP SCALE

Maps, as stated earlier, are the most prominent examples of scale. Maps represent a much larger geographical area and can be of various types depending on the features they emphasize. The area represented by a

2 ----- 4 ----- 8 ----- 10 miles
---------- 10 miles

Figure 3: Examples of differences in linear and logarithmic scales.

map can be wide-ranging, from a small room to the entire universe. Most maps use both the metric as well as U.S. measurement units.

The relationship between a specific distance on the map and its actual distance is identified by scale (also known as map scale). The distance between two points on the map can thus be easily calculated. Map scale is generally indicated in three forms—verbal, representative fraction, and bar (or graphic). The verbal scale, which is the most basic form of representation, simply gives a written description of the map-to-actual distance relationship. For example, "One inch equals one mile." This would imply that a distance of one inch on the map is equivalent to one mile (63,360 inches) on the ground.

The representative fraction (RF) scale (sometimes referred to as ratio scale) is the most flexible of all the three scales. This scale indicates that the relationship between one unit on the map is equivalent to a specific number of units on the ground. The ratio scale is flexible because the unit of measurement can be assumed to be anything (centimeters, inches, etc.). This scale is usually expressed as a fraction (and thus the name). For example, 1:50,000 implies that 1 unit on the map is equal to 50,000 units on the ground. Again, this unit could be millimeter, centimeter, inch, and so on. Simply put, the map is 1/50,000 times the size of the actual area it represents.

The concept of RF scale is often used while designing scale models of automobiles, rail, and aircrafts. The size of most automobile scale models, for instance, is 1/32 or 1:32. In other words, one unit on the scale model is equivalent to thirty-two units of the actual automobile. RF Scales are also prominently used by geographers (people who experts of geography). Many people, however, also find RF scales confusing as they do not realize what unit of measurement to use.

The bar, or graphic scale, is the most widely used type of map scale. It is merely a single line marked with distances corresponding to the ground. Given below is an illustration of different kinds of bar scale.

The first illustration in Figure 3 indicates that the distance between two adjacent points is equal to two miles. Note that this is a type of interval scale. The best way to interpret such scales is by measuring the length of any interval with a foot-ruler, and then using this length as a reference for the entire map. For example, if the length between any two adjacent points is 0.5 in (inch) on the foot-rule, a distance of 0.5 in, anywhere on the map, would indicate 2 miles on the ground. Similarly, the second illustration denotes the length of the dotted line to be equivalent to 10 miles on the ground.

Based on their scale, maps can be categorized into two types—the large scale map, and the small scale map. The large scale map shows a smaller area but in greater detail, whereas a small scale map shows a larger area in less detail. A city map would be an example of a large scale map as compared to a world map (small scale).

ARCHITECTURE

We discussed earlier how scales were used in ancient architecture. Scales (and especially map scales) are extensively used by modern day architects and interior designers. Architects always draw plans (diagrams) before starting construction work on any structure, be it a building, a house, a football stadium, or even an entire city. Such diagrams are based on the concept of map scale.

These plans give a detailed view of the entire structure in terms of size and dimension. In other words, a plan would give an architect a much better sense of the final structure. For example, a plan for a house would specify the area (length, width, and height) of every room, including the living area, bedroom, and bathroom at every floor. It would also specify details of the garage, porch, and so on. Each of these is designed with respect to a specific scale corresponding to the actual house. The main purpose of this diagram is to ascertain whether all requirements are being satisfied within the given area. For example, are there enough bedrooms, is the garage large enough, and so on.

After conceptualizing the design on paper, 3-dimensional scale models are developed. These are miniature, yet detailed, prototypes of the actual structure. They are similar to the scale models for automobiles, discussed earlier. They are a certain proportion of the actual structure.

Interior designers also develop similar diagrams of a room before designing it to get a better idea of the space (area) available to them.

In a nutshell, scale models and diagrams allow architects to visualize a structure before it is built.

WEIGHING SCALE

We are all familiar with weighing scales and have used them frequently to measure our weight, or that of other things. As the name suggests, weighing scales work on the principle of scale. A weighing scale calculates the weight of an object and displays it on a scale. The units of

the scale may be ounces (oz), pounds (lb), grams (g), or kilograms (kg). The weighing scale can be categorized as a ratio scale as it has an absolute zero and can go up to any weight depending on its type. For example, most home weighing scales indicate up to 300 lb.

Scale applications discussed till now mainly focused on the length (distance) property of an object. Weighing scales represent the mass (weight) of an object on an easy to comprehend scale. There are numerous types of weighing scales available apart from the common home scale, or bathroom scales as most people would call them. Such scales are widely used for medical, industrial (for example, weighing heavy equipment), and retail (for example, weighing food items or other groceries) purposes.

However, a majority of traditional weighing scales are now being replaced by digital scales. These simply display the weight in digital format (numbers) rather than show it against a scale. As a result, the use of scale as a weighing tool is decreasing.

THE CALENDAR

We look at the calendar every day. The calendar is merely a scale indicating the progression of time. It is one of the most universal examples of interval scale. Like an interval scale, the magnitude of intervals (years, months, days) between any two adjacent points on the calendar is the same. For example, the interval between January 1, 2001, and January 1, 2002, is the same as January 1, 2003, and January 1, 2004 (one year).

The years on a calendar (or even months, weeks, and days) can be meaningfully added or subtracted. The same is not true when you multiply or divide them. Moreover, as in the case of any interval scale, the calendar does not have an absolute zero point. The year 1 A.D. does not indicate the beginning of time. Time before this is specified as B.C.

Similarly, scale is also used in clocks.

ATMOSPHERIC PRESSURE USING BAROMETER

A number of modern-day instruments used for prediction and analysis of weather patterns include scales. One such instrument is the barometer. The barometer, like a thermometer, is an inverted glass tube dipped in mercury and sealed at the other end. It is used to measure atmospheric pressure, the weight due to the pressure (or force) of the atmosphere. As atmospheric pressure varies, the mercury in the barometer rises or dips accordingly.

The barometer consists of a scale (in inches) corresponding to the atmospheric pressure in millibars (unit for measuring atmospheric pressure). The height of the mercury on the scale would indicate the atmospheric pressure. For example, at sea level the atmospheric pressure is 1,013 millibars, and the corresponding mercury height would be 29.92 inches on the barometer scale. The atmospheric pressure at higher altitudes (height) is lower due to decreased air mass at and above the recorder.

In addition, the atmospheric pressure as indicated on the barometer scale can often be used to generally forecast weather conditions for the next twelve to sixteen hours. It is for this reason that weather experts use barometer readings in forecast reports.

An atmospheric pressure of around 1,015 millibars would indicate dry and calm weather. As the pressure increases, the temperature rises as well. In other words, higher the pressure, the sunnier are the conditions. Similarly, as the atmospheric pressure decreases, conditions usually become colder and wetter. A rapid fall in the atmospheric pressure would imply that a low pressure storm system might be approaching.

MEASURING WIND STRENGTH

Another type of instrument commonly used by weather experts is the Beaufort scale. The Beaufort scale, devised by Sir Francis Beaufort (a British admiral) in the early 1800s, is used to measure the speed of winds at sea. This instrument includes a scale of 0–12 (0–17 in some cases). Each number represents a certain strength of wind 10 m (meters) above the ground. The numbers also indicate the height of the waves in the sea, giving an idea of its state. Given below are a few observations and their implications on the weather.

A measurement of 2 (known as Force 2) would imply a light breeze blowing at a speed of 4–7 mph (miles per hour). The height of waves in the sea would be less than 0.1 meter (less than 4 inches), implying a relatively calm sea with small wavelets. A measurement of Force 6 on the scale would imply very strong breeze blowing at 25–31 mph. Such wind is capable of moving large tree branches and would make it difficult to control an umbrella if out in the open. The sea would understandably have large waves (up to almost 10 ft or 3 meters high), indicating rough conditions. Lastly, Force 12 on the scale would suggest the possibility of a hurricane. Winds would blow at enormous speeds of around 80 mph and waves in the sea can rise as high as 45 feet (approximately 14 m). Severe destruction can be caused in such cases. The devastating hurricanes that struck Florida in 2004 were measured at 12 on the Beaufort scale.

TECHNOLOGY AND IMAGING

Technology is continuously progressing. Software tools have become extremely complicated and can do

The difference in scale between an adult hand and that of an infant is clear. BETTMANN/CORBIS.

things we could not imagine a few years ago. Most software tools are based on mathematical concepts including scale. Take, for example, the car GPS (global positioning system) navigation tool. These systems help the driver in navigating from one location to another. All the driver has to do is enter the starting point as well as the destination, and the GPS would give detailed street-by-street directions on how to get to the destination. A map (much similar to a road map) is also shown on the GPS screen. The scales used by the GPS are similar to those used in printed maps.

Most architects now draw diagrams on the computer using specific software. The software tools make the architect's job easier and even faster. Scale diagrams can be effectively designed and printed. One such tool that is widely used by architects and interior designers around the world is AutoCAD.

Scale also forms an integral part of creating graphics and images, especially when using specific software tools. The building blocks of computer images are the pixels. A pixel is a specific number of blocks of color arranged in a grid. For example, a good quality 4×6 inch photograph would generally have 100 pixels per inch—400×600 pixels in total. The total number of pixels of any image is known as the resolution of that image.

The quality of an image is often measured by its resolution. In other words, the higher the resolution, the better the quality. What does scale have to do with the resolution of an image? To understand this relationship, take the above example again. If you have a 400×600 pixels computer image, you would be able to print a good quality 4×6 inch on paper. However, this does not mean that the size of the 400×600 pixels image on the computer is necessarily equal to the size of 4×6 inch image on paper. The size of the computer image (in inches) may be much smaller than the actual image printed out.

In fact, you can even print a much bigger image, say 8 × 10 inches with the same resolution (400 × 600 pixels). The quality of the printed image in this case, however would not be as good. In other words, the size of the printed image is not limited by the size of the computer image—the quality is. Using scale, you can define the relationship between pixels and inches to get a good quality image. In our example of the 4 × 6 inch image, the scale to get a good quality image can be defined as 100 pixels = 1 inch.

TOYS

Children love to play with toys. We always think of toys as a means of entertainment. However, toys also can be educational. Most toys, whether you buy them from the local Toys 'R' Us store or get them free with a kid's meal at McDonald's, are small scale representations of actual objects in the real world.

Take, for example, miniature dolls of characters from the Star Wars movies or cars from HotWheels.

THE RICHTER SCALE

The magnitude of earthquakes is measured using a numerical scale known as the Richter scale. Every earthquake, big or small, releases energy and produces shock waves of specific size within the Earth. The size of magnitude of these waves can be recorded on the Richter scale to get a better idea of how big or small the earthquake is. This scale is logarithmic. Thus, the increase of 1 unit of the scale would represent an increase in size of the shock waves (or magnitude) by 10 times. For example, an earthquake measuring 4 on the Richter scale is 10 times the magnitude of an earthquake measuring 3 on the scale.

The Richter scale starts with a unit of 1. It has no upper limit; however, it is important to note that theoretically it is not possible to have any earthquake bigger or equivalent to unit 10. An earthquake measured between 1 and 3 on the Richter scale would generally not be felt. On the other hand, an earthquake measured at 7 can be termed a massive one, capable of causing great damage up to a distance of 100 km (kilometers), or 62 mi (miles).

Tsunami, or seismic sea waves, are a series of strong ocean waves generated by the sudden displacement of large volumes of water. A very strong undersea earthquake caused a massive tsunami and killed hundreds of thousands of people in late 2004. The earthquake (ultimately measured at 9.0 magnitude on the Richter scale) created a series of waves that then radiated over thousands of square kilometers.

EXPANSE OF SCALE FROM THE SUB-ATOMIC TO THE UNIVERSE

Scale models can be used to represent parts of the solar system. Our solar system, and indeed the entire universe, are too big to comprehend. Without the use of small-scale models, it may not be possible to study them. Small-scale models are designed such that all parts of the model are in the same proportion in terms of size. In other words, the proportion of Earth's actual size to the actual size of the Sun is the same in the scale model.

To build a scale model, you must divide all sizes by a common factor. Scale models of different sizes can be built based on different common factors. However, the distances between planets and stars are so large that it may not be easy to find a common factor. For instance, the distance between Earth and the Sun is approximately ninety-three million miles, whereas the distance between Pluto and the Sun is three thousand seven hundred million miles. It is easier to make models based on distances (or sizes) measured in astronomical units (AU). 1 astronomical unit is equivalent to ninety three million miles (the same as the distance between Earth and the Sun). Consequently, creating a scale model based on the scale 1 AU = 30 inches would be much easier rather than dividing the distances by a common factor.

Over the years, scientists and researchers have used scale models to compare sizes of different objects ranging from the really large (the Sun, planets, and other stars) to those that have a more reasonable size (an elephant, or a human being) to the really minute objects that are not visible to the naked eye (an atom, an electron, or a body cell). The study of such "sub-atomic" objects is known as Nanotechnology ("nano" means really small). Simply put, the expanse of scale allows us to explore in detail, things from space and galaxies to nanotechnology.

MUSIC

Interestingly, music from the ancient period has been developed on various principles of mathematics. Scale is one of them. The most basic form of music is known as a "note." Each note corresponds to a certain frequency of sound. A series of such notes makes up music scales. The music scale is comprised of notes that are evenly spaced in terms of sound frequency. Thus, the music scale is based on the concept of interval scale.

Western music is made up of a number of major and minor scales. Each major and minor scale is comprised of seven notes. In other words, the frequency of various sounds is represented in the form of an interval scale to make a music scale.

THE METRIC SYSTEM OF MEASUREMENT

The metric system consists of units of measurement that are used to measure length, mass, or temperature. The most common units are meter (length), kilogram (mass) and Celsius (temperature). The meter, as we discussed earlier, was developed using the concept of scale. Moreover, length or mass can be measured with different units (within the metric system) related to each other by factors of 10—the basis of logarithmic scales. For example, 10 millimeters (mm) = 1 centimeter (cm), 100 centimeters = 1 meter (m), and 1,000 meters = 1 kilometer (km). This is true for units of mass as well.

The metric system is used in most countries. However, the United States and a handful of other countries still use the English system of measurement (foot, pound, Fahrenheit).

The use of the English system has always hurt the economy of the United States. It makes communication and trade with other countries difficult, and eventually affects the competitiveness of the United States. Due to this, most people within the Untied States have been encouraging the use of metric units. In fact, many organizations, including government agencies, have been using both systems of measurement.

SAMPLING

Sampling (taking small samples of a much larger quantity) is done extensively in the real world. A good example of sampling is blood tests. Doctors would take blood samples of a patient to determine his/her illness. To put it simple terms, to analyze a problem we do not study the whole object (blood, in our example) but we only take a small scale portion of it, commonly known as a sample.

Sampling is used in numerous other ways. Agriculture is another application. Farmers, around the world, grow different types of food at different places. What type of food (or crop) can be grown at a particular place depends on the nature of its soil. The soil mainly consists of four ingredients—water, air, minerals, and organic matter.

A specific crop would require the right amount of all of the above in order to grow well. Thus a farmer's job is not limited to only planting seeds and waiting for the crop to grow. The soil in different parts of his/her field must be tested (and recorded) to get a better idea of the contents of the soil. To do this, farmers take samples of soil from various parts, measure their content (using appropriate tools) and record them on a scale. The scale has an absolute zero (ratio scale). For instance, zero value of a mineral would indicate that there is no mineral in the soil sample. Similarly, if two samples (A & B) contain 2 gm (grams) and 1 gm of mineral respectively, we can say that sample A has twice the amount of minerals compared to sample B.

By recording data pertaining to the content of the soil, farmers can compare all the samples. They now know how much water and fertilizers to add and where. In other words, they can ensure that the soil contains the required amount of nutrients and water for the best possible growth of crops.

Where to Learn More

Books

Mills, Criss B. *Designing with Models: A Studio Guide to Making and Using Architectural Design Models.* Wiley, 2000.

Glazer, Evan M., and John W. McConnell. *Real Life Math: Everyday Use of Mathematical Concepts.* Greenwood Press, 2000.

Web sites

Discovery Magazine. "Size and Scale" <http://school.discovery.com/lessonplans/programs/sizeandscale> (March 14, 2005).

The IN-VSEE Project, Arizona State University. "Size & Scale: On the Relative Size of Things in Our Universe" <http://invsee.asu.edu/Modules/size&scale/unit1/unit1.htm> (March 14, 2005).

University of Washington Libraries. "How to use and understand Maps." Jan. 04, 2005. <http://www.lib.washington.edu/maps/use2.html> (Mar. 14, 2005).

Scientific Math

Overview

Numerous mathematical concepts are used for explaining complex real life situations using a scientific process. These concepts include trigonometry, calculus, rational exponents, statistical analysis, logarithms, and factoring. Mathematics, when used for scientific applications or processes, is referred to as scientific math. Such scientific mathematical concepts are widely used in real-life applications such as weather prediction, engineering, and astronomy. For example, scientific mathematical concepts can explain why bacteria multiply, thus providing a clear understanding of how to control them, and eventually benefiting from this process.

Scientific math is used in different aspects of daily life, including business-meteorology, aviation, biology, engineering, architecture, and basic sciences. The most common applications of scientific math are found in technology.

Fundamental Mathematical Concepts and Terms

The fundamental scientific math concepts cover an entire gamut of areas. Professionals such as engineers, scientists, accountants, and carpenters use scientific mathematics in different ways to manage and find solutions to improve their work. Thus, scientific math plays a vital role in various walks of life.

FUNCTIONS AND MEASUREMENTS

A function is a mathematical expression that specifies a relationship between two sets of numbers (perhaps representing physical characteristics of objects). In simple terms, it shows how a number belonging to one group can be related to a number belonging to some other group.

Take, for example, a tank of water. The total quantity of water in the tank can be expressed as a number. To estimate how much water the tank can hold, a function can be defined that presents a relationship between the shape of the tank and the volume of the tank (total quantity of water). A tank that is square would have a certain volume (quantity) of water, whereas a cylindrical tank would have a different volume. In a nutshell, the volume of the tank (and water) varies depending on the shape of the tank.

A variety of scientific processes can be explained using functions. For instance, meteorologists (people who study and predict weather) usually use relationships to predict the type of weather. There are certain factors that are vital in understanding weather patterns. Each of

U.S. Navy Blue Angels in formation. MUSEUM OF FLIGHT/CORBIS.

these factors is inter-related. Such relationships can be expressed easily in terms of functions. For example, there is a certain relationship between humidity and rain. If the weatherperson knows that the humidity is high, he or she may predict rain in the next few hours.

Functions can be simple or complex depending on the nature of the relationship. When a task becomes difficult and more factors are to be considered, a complex function may be used to explain the relationship. In the above example, if the tank is leaking, the function describing the relationship between the shape of the tank and the quantity of water would be far more complicated. This may involve the use of more elaborate mathematical concepts such as calculus and differential equations.

DISCRETE MATH

For every task, there can be two or more options. For example, a simple toss of a coin can result in a heads or tails presentation. The possibility of either a head or a tail showing up can be expressed by probability. Probability is the measurement of the likelihood of a particular event. For example, the probability of tails showing up, after a coin is tossed, is 50%. In other words, the probability of either the head or tail showing up is equal.

Workers in various disciplines, including engineers, mathematicians, marketers, government administrators, economists, biologists, or others who work with a vast collection of data, use probability and statistics. These concepts fall in an area of math commonly known as discrete mathematics. Like probability, most people also use statistics extensively in daily life activities. Statistics involves the analysis of recorded data. In other words, once a range of data is collected, it is organized and then studied to establish relationships and in turn, make more sense out of the data. For example, consider the population of a place. This data can then be organized by age or gender (male or female) and analyzed to understand various issues, such as determining the average age of the entire population, or the ratio of males to females, and so on.

Similarly, certain scientific predictions can be made from the above data using probability. For example, after determining the actual number of individuals who have received primary education from the total population of a city, the number of individuals opting for primary education can be calculated. This means that using the concepts of probability and viewing past data, one can determine future educational trends and whether the future population of this city will be educated.

It is important to note that probability is not an exact science. In other words, it allows for intelligent predictions, but can never determine an answer with absolute accuracy.

Biologists, scientists, and statisticians use the scientific concepts of discrete math (such as probability) to explain how to produce a better quality of product, such as a breakfast cereal. This process requires an understanding of how to make a better wheat plant by mixing and matching different types of wheat from across the country. Scientists use probability and statistical analysis to find matched types whose combination has the best attributes of wheat (say, a higher amount of vitamins). This combination can then be used to make breakfast cereal that has higher vitamin content.

In another example, an aeronautical engineer may study the impact of rain on the wings of a fighter plane by using another area of discrete mathematics called graphical analysis to create a model of rain drops. This eventually helps in designing a better (and more reliable) aircraft wing.

TRIGONOMETRY AND THE PYTHAGOREAN THEOREM

Mathematical principles of trigonometry, especially the Pythagorean Theorem, are used to find the height of

a mountain, the distance of an airplane from the landing airstrip, the width of a river or valley, and much more without actually measuring these manually. The Pythagorean theorem was devised by the Greek mathematician Pythagoras (569 B.C.–475 B.C.). It presents a relationship between the short sides of a right triangle and its long side. The Pythagoras equation is given as $a^2 + b^2 = c^2$ where a and b are the lengths of the short side, and c is the length of the long side of the triangle. (See Figure 1.)

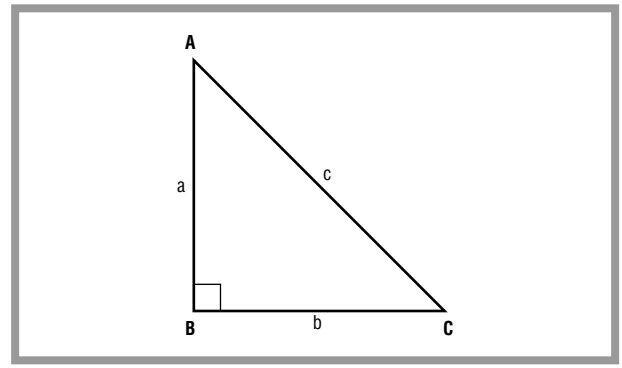


Figure 1.

Procedures that employ the Pythagorean theorem are greatly useful and far easier than manual measurement. For example, it would always be simpler to estimate the depth of a valley if it is extremely uneven and unsafe, rather than treading the terrain physically. In such cases, by taking simple measurements, the entire depth of the valley can be calculated using trigonometric concepts.

In addition, trigonometry is used for a variety of other applications. For instance, certain trigonometric concepts are used to survey distances to plan the design and development of a six-lane highway. A carpenter uses the same scientific concepts to design and build furniture. The weather department would use the Pythagorean theorem to calculate considerable data vital for predicting weather.

LOGARITHMS

Logarithms (and scales based on logarithms) are often used scientifically to simplify several processes. In mathematical terms, a logarithm is defined as the power to which a base must be raised to equal a given number. Scales that have units as log of a value instead of the value itself are known as Logarithmic scales. Simply put, the logarithmic scale represents the logarithm of the quantity rather than the quantity itself.

Logs can be expressed as exponents—$10^1 = 10$, $10^2 = 100$, $10^3 = 1{,}000$, or $10^6 = 1{,}000{,}000$ (a million). The superscripts are actually the logarithms of the final result. In other words, the log of 10 is 1, the log of 100 is 2, the log of 1,000 is 3, and so on.

A log scale generally has units such as 10, 100, 1,000, and so on, instead of 1, 2, 3, etc. The natural steps on a logarithmic scale increase in a multiplicative fashion rather than an additive or linear fashion. For example, in this case the natural steps increase by a multiple of ten ($10 \times 10 = 100$, $100 \times 10 = 1{,}000$, and so on).

Subsequently, the total length of any given road, the distance from New York to San Francisco, and even the distance from Earth to the planet Pluto can be easily compared using the log scale. In short, log scales facilitate comparison of wide ranging data. This is critical for most applications.

A geologist observing the vibrations of the earth records an extremely wide range of data (small vibrations to huge vibrations). There is a high variation in the observations, and hence it becomes much easier to present such data in terms of logarithms (and log scales).

MATRICES AND ARRAYS

A matrix is a square or a rectangular array of numbers. The numbers are represented as rows and columns. Matrices (plural of matrix) can be 2-dimensional, 3-dimensional (for example, cubical matrices), or higher-dimensional. Simply put, a 2-dimensional matrix can record different values for two characteristics of an object, whereas a 3-dimensional matrix can record different values for three characteristics of an object.

Matrices are used in several real world applications. A medical radiologist takes a patient's CT scan to show the presence of a tumor in the brain so that a surgeon can remove it. The scan of the brain uses the mathematical principle of matrix to create a 3-dimensional image. Put simply, sophisticated technology uses a matrix to convert a 2-dimensional image into a 3-dimensional one. This greatly enhances the quality of the image and makes it easier for the surgeon to pinpoint the exact location of the tumor.

Engineers use similar principles and concepts to explain the structural nature of a material. For example, using matrices, they can create a 3-dimensional view of the wings of a space shuttle. This allows them to study the characteristics (structure) of the wings in order to make them stronger and more reliable.

EQUATIONS AND GRAPHS

Scientists, engineers, and researchers use mathematical concepts of variables, expressions, or equations to show a relationship among different objects and entities. An equation is an expression made of two or more members related by an equality sign. Each member of the equation can either be a fixed number or a variable. The Pythagorean theorem described earlier is also an equation (as given below): $a^2 + b^2 = c^2$. Here, the members a, b, and c are variables representing the sides of a right angle triangle.

Equations can be used to represent different objects (and their characteristics) in the form of tables and graphs. In other words, an object can be mathematically represented in the form of an equation, or a graph, without using actual values and numbers.

The purpose of such visual representation is to understand an object (or abstract equation) better. Equations and graphs help in identifying a wide variety of patterns for the object. Moreover, complex calculations can be performed easily. In short, they help create a model to better understand a situation and to solve a problem. In addition, they also facilitate comparison between two or more objects.

A Brief History of Discovery and Development

Babylonians four thousand years ago knew of Pythagoras' theorem. As stated earlier, the Greek mathematician Pythagoras also developed this equation describes the relationship between all three sides of a right triangle. This equation has, ever since, been used extensively in architecture and carpentry. Subsequently, it was also employed in a number of other fields including the science of flight and measurements of height for various purposes.

Although ancient global civilizations knew of the importance of mathematics for various applications, it was in the eighteenth century that the Swiss mathematician Leonhard Euler (1707–1783) invented two new branches of mathematics, namely calculus of variations, and differential geometry. With these, people started recognizing the importance and value of mathematics for scientific processes.

Euler was also instrumental in pushing forward with research in number theory, which was eventually used in several scientific applications. Towards the end of the eighteenth century, Italian-born French mathematician Joseph-Louis Lagrange (1736–1813) worked on the theory of functions and equations.

Real-life Applications

WIND CHILL IN COLD WEATHER

Wind chill is a vital aspect of weather. Meteorologists provide details on wind chill along with the temperature in the winter. The reason for doing so is that a high wind chill makes a place feel colder, even though the temperature remains unchanged. In other words, a temperature of around 40°F (around 2°C) and a zero wind chill would indicate cold weather. However, the same temperature with a high wind chill would make the effects of the cold seem more efficient, although the temperature is still the same.

In simple terms, wind chill gives a measure of the discomfort caused due to a combination of the speed of wind and the air temperature at a particular time. The speed of wind does not increase or decrease temperature. Nevertheless, a high wind speed (in cold weather) always makes a person feel colder than it actually is. This is one reason why the windy city of Chicago (with normally high wind chills) has harsh winters. High wind chill can also be very dangerous, as it may result in acquiring frostbite and other problems related to extremely cold weather more quickly.

A mathematical relationship exists between the wind speed, and air temperature. It is commonly known as the wind chill index, and was devised by the National Weather Service. The equation Wind Chill Index (°F) = $35.74 + 0.6251T - 35.75(v0.16) + 0.4275T\ (v\ 0.16)$ relates the above-mentioned factors where v is the wind speed in miles per hour and T is the temperature in Fahrenheit. The unit of the wind chill index is °F. The wind chill index gives a "truer" indication of the temperature. For example, if the air temperature is 40°F and a wind is blowing at 10 miles per hour (around 16 kilometers per hour), the wind chill index is 34°F (calculated using the above formula). This implies that although the actual air temperature is 40°F, it feels like 34°F due to the wind.

Meteorologists record the necessary data and use the above equation to calculate the wind chill index. The equation makes it simpler to compute wind chill as the factor values are easily determined.

Wind chill indices are of great importance to the armed forces. Often, military personnel are posted in places with adverse weather conditions (for example, Siberia). Wind chill becomes critical in such cases. In addition, wind chill indices are also used to study the effect of wind and cold weather on other life forms.

WEATHER PREDICTION

Meteorologists predict weather patterns over the next few days or even few weeks. Weather prediction is

based on a number of factors. These include the geographical location of a place, its weather in the last few days, existing temperature, humidity, formation of clouds, wind systems, pollution, historical weather trends, air pressure, direction of the wind, and many more factors.

Meteorologists use the mathematical concept of relationship to explain the cause and effect of each of these factors (in the form of a mathematical equation). Each of these relations is then studied to develop an overall prediction. Based on these findings, weather is usually expressed in six different ways—as the temperature (in °F or °C), the humidity of the air, the type and amount of cloudiness, the types and amount of precipitation, the atmospheric pressure, and the speed and direction of the wind. These weather attributes describe future weather patterns. For example, high amounts of precipitation and cloudiness would indicate rain in the next few hours.

The weather at a particular place can also be predicted by the scientific relationship between a number of other factors, such as the amount of solar radiation, the geographic location (latitude and longitude), position of the sun, and the seasonal variation in the altitude of the sun. There are other factors including the type of rainfall, the type of clouds, and ocean currents.

In a nutshell, a large collection of factors affect weather and their relationship can be expressed in the form of an equation. These equations are also used in a host of other applications. For example, astronomers and scientists use them to predict and study weather patterns of other planets. The National Aeronautics and Space Administration (NASA) uses these findings to model spacecrafts. Similarly, climatic conditions in places having extreme weather (such as many parts of Antarctica) can be understood on the basis of data collected by studying similar relationships. Explorers can then use these weather predictions to their advantage. Meteorologists also use similar equations to understand the impact of weather on crops.

Conversions

The units of measurement around the world are based on two systems, the Metric System, and the English System. Most the countries, apart from the United States, employ the Metric system of measurement. The Metric system of measurement includes units such as centimeter, meter, kilometer, gram, kilogram, °C, and so on. The English system of measurement includes units such as foot, mile, pint, gallon, quart, °F, and so on.

There is a definite relationship between a unit of measurement in the Metric system and its counterpart in the English system. This relationship can be expressed in the form of an equation. Such equations are often used in daily life activities for converting one unit to another.

For example, the relationship between one mile and one kilometer can be shown as: 1 mile = 1.61 kilometers. Similarly, other units can be expressed in terms of mathematical equations.

Conversions based on equations and relationships are possible for any unit of measurement, even within the same measurement system. Some of the real world examples based on equations discussed earlier are also types of conversions. In other words, any equation always results in the conversion of one entity (or unit) into another.

Conversion is also used extensively in scientific processes. For example, the amount of electricity produced from thermal energy can be expressed as an equation. The calorific value of fuel can be converted into heat, and then into power using simple equations. The relationship between water, its solid form (ice), and gas form can be shown by an equation. This would state at what temperature water gets converted into ice or gas.

ESTIMATING DATA USED FOR ASSESSING WEATHER

Meteorologists must gather information about the environment to predict weather. To collect and measure data critical for determining weather patterns (temperature, pressure, precipitation, solar energy, wind speeds, and so on) they use specific tools. All these tools are set on a balloon, which is raised into the atmosphere at a particular altitude (height). The balloon is attached to a cable. Using this process, meteorologists can record data at different altitudes; the balloon is raised to a certain height and brought down to note down the readings. It is raised again to a different height, and the same data is again recorded. This process is repeated for varying heights as it helps meteorologists predict weather better. Moreover, it is also repeated on a regular basis every day. Changes in weather patterns at certain altitudes can assist in forewarning people about difficult conditions.

The height at which data is gathered is crucial. This would be equal to the length of the cable. However, measuring the length manually would be an extremely difficult

and time-consuming task (as the balloon is often raised to very high altitudes). This is where the Pythagorean theorem comes in. Using the Pythagorean theorem, the length of the cable can be easily calculated. In fact, it enables meteorologists to pre-measure lengths allowing them to simply place the balloon at different altitudes. Once the necessary data is gathered, other scientific mathematical concepts are used to forecast weather patterns.

BRIDGING CHASMS

Engineers and architects construct dams and highways in difficult terrains such as those across river gorges, and through mountain ranges. These terrains make it difficult to measure distances manually. With the help of certain instruments and mathematical concepts of trigonometry, especially the Pythagorean theorem, the distance or span of a bridge in such geographical locations can be calculated easily. (This is another application of math in scientific processes and hence Pythagoras' theorem could be considered as a scientific mathematical term.)

For example, if a bridge has to be built across a river, the engineer could use the principle of the Pythagorean theorem to calculate the width of the river (rather than measuring it manually). The theorem states that the sum of two smaller sides (squared) of a right triangle is equal to the square of the biggest side. One way of measuring the length of the river is as follows. A pole of known height is placed on one side of the river (perpendicular to the river). One side of a string is then attached to the top of the pole, whereas the other side is tied to the other side of the river (exactly at the end of the river). The string is then taken out and its length is measured. Using the length of the string and the length (height) of the pole, the length of the third side (which is the length of the river) can be calculated.

In hazardous conditions, an engineer can use this simple scientific concept of math to measure the river's width. Similar processes are also used in other areas. For example, the Pythagorean theorem is also used in space explorations. By simply studying the length of shadows, the depth of craters and height of mountains on the Moon (or a planet) can be estimated.

AVIATION AND FLIGHTS

A supersonic fighter jet is a complex machine to launch, especially from a small runway, such as the deck of an aircraft carrier. Controlling the fighter jet requires great skill and knowledge of how to fly the machine from a restricted runway. There are certain factors that control the take-off and landing of the fighter jet. The pilot applies scientific math concepts to launch an F/A-18 Hornet from an aircraft carrier. The amount of lift or force required to fly the F/A-18 Hornet can be expressed as a mathematical relationship dependent on factors such as the air density, the wind velocity, and the surface area of the wings. In order to allow the aircraft to take off, lift force must overcome gravity and equal the weight of the aircraft. This entire process can be shown as a simple math equation.

The benefit of applying this equation is that it allows the pilot to concentrate only on some key indicators of the equation to fly the plane from the deck of an aircraft carrier. The equation provides the pilot with critical data, including the ideal speed of the plane, to get the right lift for take-off or landing. The equation also shows the pilot how much time he or she has for safe landing of the plane on a shortened runway, in case there is low fuel and a large payload.

In another similar example, an athlete uses a similar relationship to assess the length and height of his or her long jump and the high jump or pole vault jump, respectively. Like fighter jets, the athlete also has a short run up but must jump as long (or high) as possible. Similar equations can thus be of great help.

Equations are also used to determine take-off and landing maneuvers for larger airplanes. For example, the pilot of a large Boeing or Airbus jet has to maneuver the plane, and approach the runway in a precise manner (for safe landing). Planning the approach ensures a smooth landing within the "touchdown zone" of the runway (this is an area on the runway that ensures that after touch down the airplane has can be smoothly and steadily brought to a halt). Pilots must sometimes execute visual approaches that vary in size, shape, and angle based on a variety of factors, such as other aircrafts on the runway, obstructions, noise abatement, and prevailing weather conditions. In other words, all these factors contribute to the safe landing of an airplane.

Pilots use mathematical concepts of relationship and equations (similar to those discussed in the case of fighter jets) in working out the approach strategy for landing the aircraft. Despite airplanes being equipped with modern technology instruments, a pilot must know and understand the relationship between various factors, to determine the distance and angle of descent, required to land the plane. As discussed earlier, such relationships can be expressed in the form of equations.

Using these landing equations, the pilot can figure the total distance required to land the plane from a particular height. This simple scientific computation enables the pilot to land safely, despite the distractions caused by differing conditions at various airports. In other words, the equation shows the effect on the landing caused by

Mathematics of Flight

Have you ever wondered how people measure the height at which an airplane (or a bird) is flying? Airplanes have advanced tools that constantly measure their altitude (height). However, these tools are based on simple mathematical principles, the same principles that can be used to measure the height of any flying object. The concept that is used here is again Pythagoras' theorem.

Airplanes continually record the distance they have traveled since take-off. With the help of instruments such as radar, it is also possible to pinpoint the exact location of the airplane, corresponding to the ground. Tools that measure altitude would then take this data and calculate the height using the Pythagorean theorem. Similarly, it is also possible to determine the exact location of an airplane if the height and total distance traveled are known, especially if for some reason the plane cannot be detected on the radar.

Mathematical principles are also used for other aspects of flight. Aeronautical engineers use concepts such as functions and equations, to find out how rain affects the wings of airplanes and eventually its flight. As raindrops are not all the same size, varying from tiny droplets to large blobs, each raindrop affects the wings differently. The concept of relationship is used here as well. The relation between the size of the raindrop and how it affects the wings (in terms of damage) can be expressed as a mathematical equation. For different sizes of raindrops, their corresponding impact on the wings is recorded (during an average rainy day, or even a storm). Using the equation, the impact for any size of the raindrop can be estimated (which would not be possible manually).

An entire range of data is then represented in the form of a graph. The purpose of doing so is similar to that discussed in the example on bacteria. After plotting all recorded values on the graph, a line can be drawn that represents the pattern in which raindrops affect airplane wings. The line can then be further extended to assess the impact of different sizes of raindrops, the ones that are not measured. Wings are central to the flight of an airplane (or any other flying object). Such equations and graphical representations allow engineers to assess the damage that can be caused during rain and thunderstorms, and in turn, build far more stable and reliable wings.

Furthermore, to study the effects of vibration on astronauts during a space shuttle launch, space engineers employ methods based on logarithms. Before launch, the vibrations felt by an astronaut inside the space shuttle are negligible (similar to those anyone would feel on the ground). However, as the space shuttle is about to be launched, the vibrations increase enormously. The magnitude of vibrations at different times during the space shuttle launch is expressed in terms of a logarithmic scale. This suggests that the vibrations increase in magnitude in multiples of ten. Another reason for using a log scale is that the magnitude of vibrations ranges drastically, from small tremors to large shuddering shakes. Consequently, this cannot be expressed on a linear scale (or by a linear equation).

a change in any one (or more) of the factors. For different airports, the magnitude (value) of the factors forming the equation may be different. For example, weather conditions at airports would vary. This would change the landing process (in terms of the distance required and the angle of descent). Thus, the equation would ensure that the pilot knows exactly what is the new distance, and angle of descent for that particular airport (for a smooth landing).

In addition, the pilot may not always have the opportunity to bring in the plane from a specific height every time. In such cases, the angle of descent would have to be modified. For these scenarios, pilots also create a graph (or chart) based on the landing equation that shows the relation between altitudes and the corresponding angle of approach. In simple terms, the angle of descent for different heights is known from this graph.

Such equations are also used in a variety of other applications. A baseball batter would use it to assess the force he requires (and the swing angle) to hit the ball for a home run. A trapeze artist uses it to define his or her swing while performing. Racecar drivers use similar equations to control the speed of their cars at sharp turns on a racing circuit. The mathematical concepts for all these remain the same.

SIMPLE CARPENTRY

Architects, designers, and carpenters need to understand dimensions of the structures they work on. Any

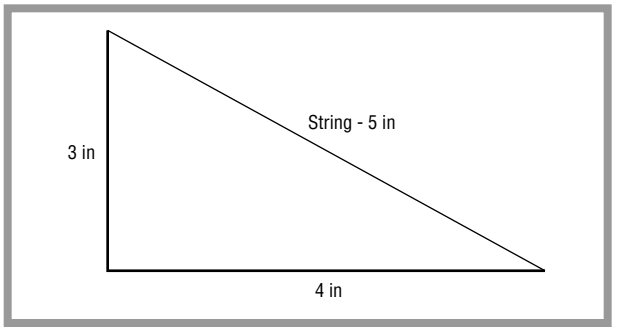


Figure 2.

construction is unique in its shape. To work out the size, shape, and dimensions of any new structure, be it a building or a simple wooden table, the Pythagorean theorem (and other trigonometric concepts) is used extensively. Although the use of mathematics in architecture is commonly known as architectural math, some of these concepts are also being used in a scientific manner, and hence can be termed as scientific math.

Continuing with the discussion on Pythagorean theorem, take for example the process of designing and creating an entertainment center for the living room. A carpenter uses the Pythagorean theorem to make sure the corners of the cabinets are at a perfect right angle. To do so, the carpenter would cut two pieces of wood (forming two sides of the cabinet). Consider that the length of one side is 3 inches, and the other is 4 inches. To join both these pieces such that they form perfect 90° angle, the carpenter would cut a string 5 inches long. As per the Pythagorean theorem, if the string fits perfectly between the free ends of both the pieces, the angle between them is an exact 90° angle. (See Figure 2.)

The same process would be repeated for the remaining two sides and a square cabinet is created.

The benefit of using the Pythagorean theorem is that the carpenter does not have to always manually measure the angle between sides, using complex tools. He or she can do this by simply measuring the corresponding lengths and sizes, a process that is far more convenient. Pythagoras' theorem can be similarly used on larger scales as well. An architect (or engineer) designing a highly technical structure, or even a model, uses this same mathematical concept.

The benefit of using trigonometry in carpentry or architecture is that the relationships between shapes and sizes hold true in most conditions. In other words, the relationship between two sides of a cabinet as defined by the Pythagorean theorem would be the same even for a much larger or more complex structure. Besides, in some cases, an architect may not need sophisticated tools. Knowledge of mathematical concepts and simple tools (such as a string in our case) can do the trick. Ancient structures around the world were built using similar methods, as they did not have advanced tools at the time.

MEDICAL IMAGING

In earlier days, in order to diagnose internal problems of a patient, doctors could only rely on x-rays that created 2-dimensional images. This would make complicated operations such as surgery rather difficult. Medical imaging, over the years, has progressed immensely. Newer technologies that produce 3-dimensional images have become extremely common. These technologies that greatly facilitate complex operations are based on mathematical concepts.

A radiologist undertakes imaging of the human body to find any growth, say in the brain, using computer tomography (CT) or nuclear magnetic resonance (NMR). An explanation of these technologies is not within the scope of this article. For our understanding, these are imaging methods that use mathematical principles of algebra and matrices to create 3-dimensional images.

CT measures the length of x-ray beams passing through a part of a body, from hundreds of different angles. Subsequently, based on the evidence of these measurements, computer software is able to reconstruct 3-dimensional pictures of the body's interior. In doing so, the software uses matrices to define dimensions of small portions from the body. In other words, the body (or a part of it) is considered as a number of smaller parts. The dimension of each part is defined using a 3-dimensional matrix. The 3-dimensional matrices for all parts are then joined together to get a complete image. Sonography is another technology that is based on similar concepts.

As stated earlier, the benefits of 3-dimensional imaging are numerous. Doctors can pinpoint the exact location of a problem area and perform surgery with greater effectiveness. Simply put, doctors can see the inside of any part of the body (as if they are the actual thing itself) and diagnose health-related problems far more efficiently. Three-dimensional imaging is also used in a range of other applications. This includes architecture, aviation, automobile engineering, computer games, and much more.

ROCKET LAUNCH

Rocket scientists are always looking for cheaper and more effective ways of launching a rocket. The space agency NASA launches its space shuttles from Florida, and with a reason. According to the Coriolis force, a scientific

Bacterial Division and Replication

Medical scientists, biologists, and health officials constantly study the impact of various diseases on humans and other living beings in order to seek better cures. To study these diseases, these scientists must first understand how bacteria multiplies and at what speed. Bacteria are small cells (living organisms) that cause many diseases. Once inside another body, bacteria multiply quite rapidly. Each single-celled bacterium divides to form two or more bacteria cells. Each of these then split into two or more bacteria cells. The process, known as binary cell fission, is efficient at causing tremendous bacterial growth.

The time that it takes one bacterium to split into two (or more) cells is known as generation time. Generation time varies greatly among different species of bacteria. For example, certain bacteria such as *Escherichia coli*, which causes severe diarrhea, takes only about twenty minutes, whereas *Mycobacterium tuberculosis*, the bacterium responsible for tuberculosis, would take as much as twenty-four hours.

Cell fission follows certain mathematical principles. Consequently, the entire process can be expressed through algebra and basic calculus. Scientists use these mathematical concepts to develop a model for understanding the behavior of bacterial division and replication. For example, they would be able to figure out up to what number different bacteria can grow within one hour, and eventually an entire day. Most bacteria grow proportionately. In other words, the rate of replication is proportional to the population of existing bacteria. This has been established by studying the growth of bacteria in different environments, at different intervals of time.

Based on this fact, relationship models (or equations) are developed between existing bacteria population and time. Such relationship models ensure a better understanding of bacterial growth, which is extremely vital to the progress of medical science. Furthermore, scientists also study the impact of the bacterial growth through a visual graph, known as the exponential graph. The population growth trend for bacteria is an exponential curve. Each generation doubles in number. For instance, the process starts with one bacterium that replicates itself to two bacteria, two grow into four, eight, and so on. The numbers (quantity of bacteria cells) can be plotted on a graph against time. A single straight line that connects most (if not all) of these points would represent the growth trend of the bacteria. The purpose of such models is to estimate the growth after a certain time, by way of extrapolation. In other words, the line on the graph can be drawn further to estimate the growth of bacteria after subsequent intervals of time. Thus, instead of actually measuring the growth every hour through experiments, scientists can simply predict it.

Similar models can also explain the growth of a particular disease among living beings. Most contagious diseases (diseases that are spread through contact) affect living beings exponentially, at least early on in an epidemic. For example, initially one person may be infected, and after a certain time two would be infected, and so on. Exponential graphs can be used here to predict the number of people affected after a certain period of time. This helps government officials control the problem, especially in cases of an epidemic.

Similarly, the utility department can use such relationships towards controlling water borne diseases. Similar principles are also employed in handling radioactive materials in medicine, studying the impact of their exposure in space, or in case of a nuclear accident.

principle (based on mathematics) developed by French Mathematician Gustave Coriolis (1792–1843), as Earth's axis of rotation at the poles is nearly vertical to the horizon, any still object in the sky would spin horizontally. In contrast, at the equator Earth's axis are nearly horizontal. Subsequently, still objects in the sky would move vertically. Thus, objects here would get a vertical boost.

In other words, if a space shuttle was launched nearer the equator it would get a vertical boost. Consequently, because of this boost, less fuel is utilized. In fact, NASA saves considerably on the cost of a launch dollars because of this strategy. In a nutshell, the science behind rocket launch is based on a principle that presents a mathematical relationship between Earth's axis of rotation at a particular place, and the corresponding direction and speed at which an independent object moves at that place.

The same principle can be used to explain why most airplanes that fly around the world fly near the poles rather than the equator. At the poles, an object spins horizontally. In other words, its horizontal speed would be higher. Subsequently, the speed of airplanes near the poles would be higher.

Key Terms

Equation: A mathematical statement including an equals sign.

Function: A mathematical relationship between two sets of real numbers. These sets of numbers are related to each other by a rule that assigns each value from one set to exactly one value in the other set. The standard notation for a function, $y = f(x)$, developed in the 18th century, is read "y equals f of x." Other representations of functions include graphs and tables. Functions are classified by the types of rules which govern their relationships.

Logarithm: The power to which a base number, usually 10, has to be raised to in order to produce a specific number.

Matrix: A rectangular array of variables or numbers, often shown with square brackets enclosing the array. Here "rectangular" means composed of columns of equal length, not two-dimensional. A matrix equation can represent a system of linear equations.

SHIPS

The design of ships and submarines involves extensive use of a scientific principle known as the Archimedes principle. According to this principle, floating objects (or even objects that are fully or partially submerged in a fluid) displace a certain amount of fluid. Due to this, the object feels a certain up-thrust. The magnitude of the up-thrust is equal to the amount of fluid displaced. Note that this principle is a mathematical equation based on the relationship between amount of fluid displaced, and the up-thrust experienced by the object. Ship architects and engineers use this principle to assess how a ship would lie in the water before it is launched. In other words, the Archimedes principle helps in designing the ship such that its movement and position is ideal once it is in water. The same principle is also used for submarines.

GENETICS AND MATHEMATICS

Doctors and scientists perform studies to predict what characteristics a child inherits from his or her father and mother. Such scientific studies are based on the concept of probability. Consider that the father and mother have brown eyes. However, it is possible that the child does not have brown eyes. The reason for this is that every characteristic of the human body (in this case color of the eyes) has two genes (cells that possess the characteristic). One gene is dominant, whereas the other is recessive. In this case, both father and mother would have one gene that is responsible for the brown color. Simultaneously, they would also have another gene (responsible for the eye color) that would have some other color. A child is likely to inherit the dominant genes, which may be the non-brown ones, and hence not have brown eyes at all.

Using the principles of probability, scientists can figure out the characteristics a child is most likely to have. The scientist must also have complete information on the genes of the parents. Thus, characteristics such as the color of the skin, color of the eyes, facial features, build and physique, and much more can be predicted. It is important to note that scientists and doctors can only state characteristics that a child is most likely to inherit.

Scientists also use these principles to create genetically modified plants and animals. For example, scientists can genetically modify a cow so that she gives birth to a calf that ends up giving more milk. Similarly, different breeds of animals are also genetically modified so that they give birth to offspring of mixed breeds.

EARTHQUAKES AND LOGARITHMS

A common example of a logarithm scale (scale based on log values) is the Richter scale for measuring magnitude of earthquakes. During an earthquake, an enormous amount of energy (in the form of heat) is released from the surface. The magnitude of the earthquake depends on the amount of energy released. The magnitude is shown on the Richter scale.

The relationship between each step on the scale (magnitude of the earthquake) and the corresponding amount of energy released by the earthquake can be explained by an algebraic equation. The Richter scale is a log scale; the difference in magnitude (in terms of the energy released) between two consecutive steps on the scale is ten-fold. For example, the amount of energy released by a 6.0-magnitude earthquake on the Richter

scale has ten times more than the energy released by a 5.0-magitude earthquake on the same scale.

Potential Applications

GENETICS

The use of probability in genetics is fast increasing. Genetic technology is being used in agriculture and forestry to improve plants, increase disease resistance and the yield of crops, and adapt non-native crops to a new environment for specific benefits. This has improved the quality and quantity of food production in many parts of the world.

Animal breeders have been using genetic principles for many years to develop characteristics within a species that they feel are desirable. There are areas where genetics can be used; all of these are based on the principles of probability. However, the maximum benefit would be in the field of medicine. Genetic modification of certain bacteria has helped find better cures for many diseases, and holds true potential.

Where to Learn More

Books

Hazewinkel, Michael. *Encyclopedia of Mathematics.* Dordrecht, Netherlands: Kluwer Academic Publishers, 2000.

Owen, George E. *Fundamentals of Scientific Mathematics.* Mineola, NY: Dover Publications, 2003.

Web sites

"Physical Effects of the Earth's Rotation (Coriolis Effects)" <http://cseligman.com/text/planets/coriolis.htm> (April 9, 2005).

"Scientific Mathematics/Mathematics as a Science" <http://huizen.dto.tudelft.nl/deBruijn/grondig/science.htm> (April 9, 2005).

Scientific Notation

Overview

When dealing with very small or very large numbers, such as within the scientific fields of biology, chemistry, engineering, mathematics, and physics, professional men and women use scientific notation as an efficient way to read and write such numbers. The method of scientific notation uses the significant digits of a number and multiplies it by specific integral powers of ten. This method of notation is an easy way to read and write numbers so that the resulting representation makes more sense. It is also quicker to perform various mathematical operations such as addition and multiplication when large and small numbers are changed to scientific notation.

Scientific notation is used when working with the numerous sizes and time frames often found in the sciences, such as the very large distances between stars and the very small diameters of atoms and molecules.

Fundamental Mathematical Concepts and Terms

When using scientific notation, the expression of a number n is represented as $n = a \times 10^p$, where the variable a (generally representing a number between 1 and 10) is multiplied by an integer power (p) of 10. A power of 10^p is 10 multiplied by itself a specified number of times (p). For example, the number 1 would be written as 1×10^0 where $10^0 = 1$; 10 would be written as 1×10^2 where $10 \times 10 = 100$; and 1,000 would be written as 1×10^3 where $10 \times 10 \times 10 = 1{,}000$. As a specific example, when n = 71,000, the number n can be written in scientific notation as 7.1×10^4, where a = 7.1 and p = 4.

Writing numbers in scientific notation allows scientists to reduce, and often times eliminate, many zeros while indicating that zeros are still significant. The number 71,000, as shown above, is the same as 7.1 multiplied by 10,000 (10^4) and is written 7.1×10^4 in scientific notation. In scientific notation, numbers that are smaller than one will contain negative exponents. The number 0.00523, for example, is the same as 5.23 times 0.001 (10^{-3}) and is written 5.23×10^{-3}. (The term 10^{-3} means that 1 is divided by ($10 \times 10 \times 10$).)

To convert a large or small number to scientific notation, move the decimal point in the number to the right of the first nonzero digit. For example, within the number 45,630,000.00, move the decimal point seven places to the left so that it is positioned to the right of 4, the first nonzero digit. (Remember that a decimal point is implied at the end of every whole number, even when it is not

written; thus, 33 = 33.0.) Then, indicate the movement by multiplying by a power of 10 that shows the exact number of positions moved. In this case, since the decimal point was moved seven positions to the left, 4.653 would be multiplied by 10^7, or 4.653×10^7. If the decimal point is moved to the left (as in the above example), the exponent p in $a \times 10^p$ is positive (+). If the decimal point is moved rightward, then the exponent p is negative (−). In this second case, the number 0.0000376 would be written as 3.76×10^{-5}.

The process is reversed when changing a number from scientific notation to regular notation. That is, move the decimal point the same number of places as the value of the exponent and then move the decimal point to the right if the exponent is positive or to the left if it is negative. Finally, add zeros if necessary.

Scientific notation is also important because it provides a clear indication of the number of significant digits within a number or calculation. For example, if a truck weighs 4,007 pounds (accurate to the nearest pound) then both 4,007 and 4.007×10^3 give an accurate measurement. However, if a truck's weight is 4,000 pounds then it is not (necessarily) apparent that this weight is accurate to the nearest single pound because it might be rounded to the nearest ten pounds. Scientific notation, on the other hand, shows that all four digits are significant. That is, when the truck weight is shown as 4.000×10^3 pounds, the extra (significant) zeros to the right of the decimal point show that the precision of the measurement is down to the single pound.

A Brief History of Discovery and Development

The publishers of the Oxford English Dictionary are interested when a new word or term shows up in print for the first time. The first recorded use of the term scientific notation appeared in the third edition of the *New International Dictionary of the English Language*, which was published in 1961. Because scientific notation did not appear in the second edition of that dictionary, which was published in 1934, language experts widely assume that the term was probably invented sometime during the decades of the 1940s or 1950s, and it is also assumed that the term gained widespread usage during the 1960s.

In 1963, the term was used inside the article "Digital Computer Technology and Design" that was part of the Oxford English Dictionary. Within this article, scientific notation referred to any number of the form: a first number times a second number raised to a third number. Since scientific notation came from a computer science reference book, the term is likely to have been regularly used by the pioneering computer users who were already buying electronic calculators and experimenting with simple computers. It is assumed that computer enthusiasts wanted a specific way to describe how a number is stored in a computer because at that time there was a big difference in how integers and fractional numbers were stored. (By the way, it is assumed that mathematicians, physicists, and engineers did not invent the term because they were already using the term exponential notation as an alternate form for scientific notation.) In 1973, scientific notation was defined in an introductory textbook on computer science and two years later appeared in the *Physics Bulletin* as a feature contained on calculators. By this time, the term scientific notation had spread from the computer science community out into the community of physicists and other physical scientists.

The modern meaning of the term scientific notation has changed from its original meaning. In the 1960s the meaning of scientific notation referred to any number of the form "first number times second number raised to third number." In modern usage of scientific notation, the second number is always 10, while the more general term exponential notation is used when this second number is any numerical value.

Real-life Applications

MATHEMATICAL OPERATIONS

One of the most obvious reasons to use scientific notation is when adding, subtracting, multiplying, and dividing very large and very small numbers. Adding two or more numbers with scientific notation involves converting all of the numbers to the same power of 10 and then adding the digit terms of the numbers. (When measurements are added or subtracted, the accuracy of the answer is no greater than the least accurate measurement.) For example, adding 5.045×10^{-6} and 2.65×10^{-4} involves: $5.045 \times 10^{-6} + 265 \times 10^{-6} = 270.045 \times 10^{-6} = 2.70 \times 10^{-4}$. When subtracting two or more numbers with scientific notation, convert all of the numbers to the same power of 10 (as with addition) and then subtract the digit terms of the numbers. For example, $7.99 \times 10^5 - 4.534 \times 10^3 = 7.99 \times 10^5 - 0.04534 \times 10^5 = 7.94466 \times 10^5 = 7.94 \times 10^5$.

Multiplying two numbers with scientific notation involves using the rules of exponents: $10^m \times 10^n = 10^{m+n}$. (When measurements are multiplied or divided, the answer contains no more significant figures than the least accurate measurement.) As an example, multiplying

46,850 and 0.0000417 with the use of scientific notation results in $(4.685 \times 10^3)(4.17 \times 10^{-5}) = (4.685)(4.17) \times 10^3 \times 10^{-5} = 19.53645 \times 10^3 + (^{-5}) = 19.54 \times 10^{-2} = 1.95 \times 10^{-1}$. Dividing (/) one number such as 650,000 by another number 25,000,000 results in: $6.5 \times 10^5 / 2.5 \times 10^7 = 6.5 / 2.5 \times 10^5 \times 10^{-7} = 2.6 \times 10^{-2}$. Generally, from between one-hundredth (1/100) to 100, scientific notation is not usually needed, but on either side of this range, it is useful to apply scientific notation.

CHEMISTRY

In chemistry—which involves the study of the interactions between matter and energy and the composition of matter itself—the calculation of the parts of matter such as the dimensions of atoms uses very small numbers and thus needs the application of scientific notation. For example, the weight of a single atom of hydrogen is better expressed as 1.7×10^{-24} grams, rather than 0.000000000000000000000017 grams.

One of the fundamental laws of chemistry is Avogadro's Law, which expresses the fact that under a closed, theoretical environment of pressure and temperature, identical volumes of gases contain an equal amount of molecules. The law helped in the early development of chemistry, but the number itself (Avogadro's number) was not calculated until the last half of the nineteenth century due to limitations in technology. Avogadro's number is usually stated as the number of molecules existing in one mole, or gram molecular weight, of a substance; and is formally defined as the number of atoms in 12 grams of the element carbon-12. The number of molecules in one mole has been scientifically determined to be approximately 6.0221367×10^{23} molecules, a number that is obviously easier to read under scientific notation than with many zeros.

ELECTRICAL CIRCUITS

Physics is the study of all the physical events that take place in the universe. Physicists use mathematical equations to describe and predict physical events, and because there is so much variety in the universe, physicists encounter very large and very small numbers within their observations, theoretical calculations, and experimentation. Because of that fact, all the many divisions within physics, such as astronomy electromagnetism, mechanics, optics, quantum mechanics, and thermodynamics, use scientific notation.

In electromagnetism, for example, a scientist might consider the number of electrons passing through a point in an electrical circuit of one ampere every second. Scientific notation would be helpful during such calculations because one ampere contains about 6,250,000,000,000,000,000 electrons per second; that is, approximately 6.25×10^{18} electrons per second. In this case, the advantages of scientific notation are obvious: the number is not as cumbersome when written on paper, and the significant digits are easy to identify. Both advantages are important in nearly all physics calculations.

LIGHT YEARS, THE SPEED OF LIGHT, AND ASTRONOMY

Astronomy is the study of the universe, including all materials such as celestial bodies, dust, and gases within it. Work within astronomy includes theories and observations about the solar system, the stars, the galaxies, and the general structure of space itself. Much of the research performed in astronomy involves very large and very small numbers so that scientific notation is regularly used as an important way to handle such numbers. For example, one of the largest objects visible to the naked eye is the Andromeda galaxy, with a diameter of over 1.0×10^5 (100,000) light years. A light year is the distance that light travels in one year, or about 5.9×10^{12} miles (about 9.5 trillion kilometers). The Andromeda galaxy appears large even though it is over 2.0×10^6 (two million) light years away from the Earth.

Astronomers use scientific notation when measuring the distance between stars because these distances can be as small as a few light years to hundreds of light years or more—but in any case are very large numbers. The speed of light is approximately 3×10^8 meters per second (or about 1.86×10^5 miles per second). Even the simple calculation concerning the time it takes light to travel from the Sun to the Earth involves two very large numbers. This calculation can be simplified with the use of scientific notation by knowing the distance between the Earth and Sun is about 1.5×10^{11} meters (m) and the speed of light to be 3×10^8 meters per second (m/s). Dividing the first number by the second results in (1.5 m) / (3 m/s) $\times 10^{11} \times 10^{-8} = 0.5 \times 10^3 = 500$ seconds, or about 8.3 minutes.

Astronomers also use scientific notation when examining electromagnetic radiation from all of the wavelengths emitted by celestial bodies. The most obvious radiation with respect to humans is visible light, which is the only radiation that humans can see with their eyes. However, visible light is only a small part of the electromagnetic spectrum that includes radio waves, microwaves, infrared light, visible light, ultraviolet light, x-rays, and gamma rays. Radio waves, which include waves used to provide sound in AM and FM radio, are the longest type of radiation, with a wavelength

of about 1×10^8 meters. The shortest type of radiation is gamma waves, with a wavelength of about 1×10^{-16} meters. Gamma waves are emitted by such objects as radioactive isotopes and in some nuclear reactions, both created by mankind and occurring within the center of stars.

Professional astronomers use powerful telescopes, computers, and instruments while performing their jobs. Most work includes, first, observing astronomical bodies by using telescopes and instruments to collect relevant information. Astronomers, secondly, analyze the resulting images and data. Computational astronomy is one way that astronomers use computers and scientific notation to simulate and analyze astronomical events. Examples of events that are simulated by computers include the massive explosions of stars as they end their lives to make way for supernovas and the creation of the earliest galaxies within the universe. Thirdly, they compare their results with existing theories to determine whether their observations coincide with what theories predict, or whether the theories can be improved or, in some cases, replaced. Some astronomers work only on observation and analysis, and others work primarily on developing new theories, but in all cases the men and women within the field of astronomy use scientific notation in order to do their very difficult and complicated work.

COSMOLOGY

Cosmology, a branch of astronomy, is the study of the origins of the universe. It includes the Big Bang theory, which is the currently accepted explanation of the beginning of the universe. The theory proposes that the universe was once extremely dense and hot. Then a cosmic explosion called the big bang happened about 1.37×10^{10}, or approximately 13.7 billion, years ago, and the universe has ever since been expanding and cooling. Cosmologists best understand the universe from about one hundredth of a second after the big bang through to the present day. However, particle cosmologists attempt to describe the state of the universe that occurred only 1.0×10^{-11} seconds after the big bang—information that is hard to verify. For that reason, sophisticated computer models along with scientific notation are used in order to make predictions about unknown characteristics from the few facts known—a process called extrapolation.

Besides working on the early beginnings of the universe, some cosmologists work with quantum cosmology in order to study the origin of the universe itself. Because of the tiny and huge numbers involved, cosmologists use scientific notation within their research. This study is an attempt to characterize processes at the earliest times of the universe, that of the Planck epoch at 1.0×10^{-43} seconds after the big bang. It is widely accepted that from the instant of the big bang to about 10^{-32} seconds later, the universe expanded much more rapidly than it did later—to about 10^{50} times its original size. At the Planck epoch, the universe was extremely hot in temperature. In fact, cosmologists do not even talk in terms of familiar temperature units such as degrees Kelvin, Celsius, or Fahrenheit, but use gigaelectron volts (GeV) when dealing with such very hot temperatures. At the Planck epoch, the temperature of the universe is believed to be 10^{19} GeV, which is equivalent to about 1.0×10^{32} degrees Kelvin.

ENGINEERING

When engineers use scientific notation they call it engineering notation because the powers of ten are limited to multiples of three. For instance, electronic multimeters are set up in ranges that accommodate engineering notation. A reading of 3.06×10^{-5} amperes would not be valid (because 5 is not a multiple of 3) but with the use of engineering notation the value would be converted to 30.6×10^{-6} amperes (amps) and represented as 30.6 microamps, where micro stands for one millionth of an ampere. The prefixes associated with engineering notation include (in the positive) 10^3 = kilo, 10^6 = mega, 10^9 = giga, 10^{12} = tera; and (in the negative) 10^{-3} = milli, 10^{-6} = micro, 10^{-9} = nano, 10^{-12} = pico.

COMPUTER SCIENCE

Computer science involves the engineering, experimentation, and theory that goes into the design, production, and use of computers. Writing out very large and very small numbers can be tedious and cause mistakes, which is one reason why in early computers these large and small numbers were often written out with scientific notation. However, when inputting such large and small numbers into a computer with scientific notation, another problem arises. A number such as 1.4×10^5 was not easy to input into early computers because the times ($\times$) symbol is different from the letter "x," and most of the computers of those early years did not have a way to indicate superscripts. So, when computer languages were first developed, an alternative way of writing scientific notation was developed, the exponential notation. The "$\times$ 10" (or times 10) was replaced with the capital letter "E" and the exponent itself was written without the superscript. Thus, the value of 1.4×10^5 was written as: 1.3E5 (some other equivalent representations include +1.4E + 05, 1.4E + 05, and 1.4000E05). Because computers in the twenty-first century are used in every conceivable field of science, business, and everyday life, the

inputting of very large and very small numbers is now an easy task.

MEDICINE

The process of diagnosing, treating, and preventing illnesses, diseases, and injuries is called medicine. Although practicing doctors and other similar health care professionals rarely use scientific notation as part of their daily routine, medical researchers and scientists often use scientific notation in their search for new medical knowledge and technology in such areas as drugs, medical treatments, and equipment and devices.

Controlled clinical trials are a method used by medical professionals to decide whether new drugs and treatments are safe. In a controlled clinical trial, a group of patients, normally called the treatment group, receives a new drug or treatment. Another group, referred to as the control group, is given a placebo (an inactive drug) or a currently accepted method of treatment. Over an appropriate period of time, researchers compare the two groups as to their overall reactions. The resulting data is collected and analyzed with statistical techniques, which includes scientific notation, to determine if the new treatment is better than standard treatments or no treatment at all. For instance, in one study, volunteers might receive a one 1-milliliter ($1 \times 10^{-3} -$ mL) injection of an experimental vaccine at a dosage of 1×10^{10} units. Due to the extremely small amount of vaccine given to the volunteers, scientific notation would be used to analyze the resulting data, and to ultimately determine the safety of the new vaccine. Amounts of such medical materials as bodily fluids, drugs, DNA samples, and plasma and blood all potentially need to be measured in terms of scientific notations in order to be efficiently researched and analyzed by the medical community.

ENVIRONMENTAL SCIENCE

The study of the environment involves dealing with all of the external factors such as other living organisms and nonliving factors like ocean currents, rainfall, and temperature affecting an organism. Environmental scientists study the long-term consequences of human actions on the Earth's environment and other smaller environments. During their studies, these scientists are confronted with many large and small numbers. For instance, there are 3.34×10^{22} (or 33,400,000,000,000,000,000, 000) molecules in one gram of water, the life providing material of all organisms on Earth.

Environmental scientists are likely to measure the average volume of river water flowing into a particular ocean, which may commonly reach values of 1×10^9 cubic meters per year. For example, at New Orleans, Louisiana, the average flow rate of the Mississippi River, one of the principal freight transportation arteries in North America, is 6×10^5 cubic feet per second, which relates into 1.9×10^{13} cubic feet per year (5.4×10^{11} cubic meters per year). Because the Mississippi River is so important to the health of the United States, it is necessary for environmental scientists to study the river's overall condition. Because so many of the river's statistics are large numbers, scientific notation is regularly used to analyze the very large numbers that describe the Mississippi River with regards to transportation, farming, fishing, and the general environmental conditions of the areas surrounding the river.

GEOLOGIC TIME SCALE AND GEOLOGY

The study of the history, features, and the processes acting upon the planet Earth is called geology. A specific type of calendar is used by geologists in order to find out (for example) how long ago a dinosaur lived or why a volcano was formed. Such a calendar, which is able to go back millions of years into Earth history, is called the geologic time scale. It begins when the Earth was first formed, about 4.6×10^9, or 4.6 billion, years ago, and continues up to the present. Instead of months and days, the geologic time scale divides Earth's history into: (1) eons (the longest unit of geologic time comprising several eras), (2) eras (the second longest unit of geologic time comprising several periods), and (3) periods (a third longest unit of geologic time, shorter than an eon or an era).

Scientific notation is critical to the proper use of the geologic time scale because of the large numbers involved. For instance, the Hadean eon occurred about 4.6×10^9 years ago, at the time when the Earth was formed, while the Proterozoic eon occurred about 2.2×10^9 years ago, at the time when the mechanics of plate tectonics began to slow down and operate much like it does today. During the Jurassic period, the second division of the Mesozoic era which occurred about 2.06×10^8 years ago, reptiles were the dominant form of animal life, having adapted to life in the air, in the sea, and on the land.

FORENSIC SCIENCE

Forensic science—the application of science to law—uses advanced technologies to uncover scientific evidence in a variety of fields. There are many subspecialties within the field of forensic science including anthropology, biology, chemistry, pathology, odontology, toxicology, psychiatry, and physics. Forensic scientists in each subspecialty use scientific notation in their own way to perform the

science necessary within their specific jobs. In fact, the amount of digital evidence that forensic scientists collect each year and store in data storage devices is greatly increasing. Many digital evidence computer programs are searching terabytes of data each year, where one terabyte is one thousand billion, or 1×10^{12}, characters.

Recent technological developments easily permit scientists to analyze the deoxyribonucleic acid (DNA) of forensic evidence in order to determine whether it came from a victim or a suspected criminal. During the last quarter of the twentieth century, its use in forensic science has dramatically helped to solve ever increasingly complicated crimes. In fact, the high-technology process known as polymerase chain reaction (PCR) is an impressive technique that quickly multiples very small samples of DNA into much larger samples and results in the use of scientific notation when very small numbers are converted into very large numbers. Repeated cycles of replication (multiplication) involve the heating and cooling of a DNA sample within a solution of heat-resistant enzymes. This action results in a particular DNA sequence being multiplied at a rapid rate. Within several hours, a tiny sample, for instance 1 nanogram (or 1×10^{-9} grams) of DNA, would have been increased by about a million times, to give a milligram of sample material, enough material for many DNA tests by numerous laboratories.

Real-Life Math and Audio Engineers

Audio engineers make audio amplifiers such as those found commonly in radios and television sets. The word audio refers to signals with information in the frequency range that is audible by humans, which involves very large numbers calculable with scientific notation. An audio amplifier consists of an electrical circuit manufactured to increase the current, power, or voltage of an applied signal, which is then converted to sound. For example, electromagnetic signals between 300 hertz and 3,000 hertz are called audio-band electromagnetic signals. Generally, audio signals operate at frequencies below 20,000 hertz, or 20 kilohertz-where 1 kilohertz equals 1,000 (10^3) cycles per second-but can operate up to 100 kilohertz (or 100,000 (10^5) cycles per second).

ELECTRONICS

Electronic engineering, the largest field within engineering, is concerned with the application, design, development, and manufacture of devices and systems that use electrical power. While working in the field, electronic engineers encounter many very large and very small numbers that would be quite inconvenient to write out with traditional notation. For example, a capacitor might have a value of 0.000001 farad or a resistor a value of 150,000 ohms. Because these numbers are inconvenient to write, it becomes much easier to use 1×10^6 Farad or 1 micro-Farad and 1.5×10^5 ohms or 150 kilo-ohms. Another familiar measure used in the electronics field is the coulomb, which stands for the quantity of electrical charge. One coulomb equals the amount of electrical charge carried by 6.25×10^{18} electrons.

ABSOLUTE DATING

Anthropology is the study of all aspects involving the ways and means that humans live. Anthropologists study such topics as: what people think about, how they react to their environments, and the reasons that humans evolved over time. Archaeologists use specialized methods and tools for the excavation and recording of recovered remains of ancient peoples and their artifacts. They use a variety of dating methods—all using scientific notation to achieve reliable results—which involve various scientific analyses to uncover the characteristics of materials buried for thousands, even millions of years. One of these dating methods is called absolute dating, which determines the age of a material with respect to a particular time scale that involves very large numbers.

EARTH SCIENCE

Earth science, as the name suggests, is the study of the Earth. Because Earth science deals with many very large and very small numbers, it is essential that earth scientists use a form of shorthand to represent such numbers. As a result, scientific notation is used, for example, to represent very large numbers such as the mass of the Earth as 6.000×10^{24} kilograms, rather than as 6,000,000,000,000,000,000,000 kilograms, and the average circumference of the Earth as 4.0074×10^7 meters, rather than 40,074,000 meters. Scientific notation is also used to represent very small numbers within earth science such the concentration of gold in seawater as 5×10^{-8} grams per liter rather than the unwieldy 0.00000005 grams per liter. In each case, it would be both confusing and requiring of a great deal of space to continually use the longer version.

Key Terms

Anthropology: The study of humankind.

Exponent: Also referred to as a power, a symbol written above and to the right of a quantity to indicate how many times the quantity is multiplied by itself.

Potential applications

PROTEINS AND BIOLOGY

Biology is the study of life, and its data is usually acquired through measurements of very small and very large values of mass, volume, length, temperature, pressure, and pH, which again necessitates the need for scientific notation. For example, the basic length scale used to describe any type of molecule is a nanometer, where 1 nanometer $= 1 \times 10^{-9}$ meters. Currently, biological data is subdivided into sections based on the cell, the molecule, the organism, and the population. Using an example from molecules, when compared to a water molecule, a protein molecule is gigantic with a typical mass of about 1×10^{-22} kilograms. Made up of thousands of atoms (mostly atoms of carbon, hydrogen, oxygen, and nitrogen), proteins serve numerous purposes such as a constituent of bones and tendons; an ingredient within red blood cells; a part of oxygen in the lungs; a material in the hair and skin, and an aid in the digestion of food. Discovering how atoms are arranged in a protein molecule is one of the most challenging research projects in the biological sciences. With an estimated 30,000 different proteins in the human body, only about two percent have been adequately described, which provides ample need for future research into the discovery of these descriptions, and along with it the need for calculations with scientific notation.

NANOTECHNOLOGY

Nanotechnology is a relatively new science that involves the creation and use of materials and devices at extremely small sizes. These materials and devices are generally in the (nanoscale) range of 1 to 100 nanometers, where one nanometer is equal to one-billionth of a meter (0.000000001, or 1×10^{-9}, meter), which is about 50,000 times smaller than the diameter of a single length of human hair. Scientists refer to the materials at the nanolevel as nanomaterials or nanocrystals. The transmission electron microscope, a pioneering nanotechnology invention, is already a popular instrument for visualizing individual atoms within semiconductor nanocrystals. This instrument and other such breakthroughs have already applied nanotechnology, but future research and development will hold the key to nanotechnology's major impacts in such fields as energy conservation, medicine, environmental protection, electronics, computers, and world defense.

Where to Learn More

Books

Bluman, Allan G. *Mathematics in Our World.* Boston, MA: McGraw-Hill-Higher Education, 2005.

Florian, Cajori. *A History of Mathematical Notations.* New York: Dover Publications, 1993.

Tussy, Alan S., and R. David Gustafson. *Basic Mathematics for College Students,* 2nd Ed. Pacific Grove, CA: Brooks/Cole Thomson Learning, 2002.

Web sites

Bagenal, Fran. Atlas Project, University of Colorado at Boulder. "The Scientific Notation." Mathematical Tools for Astronomy.<http://dosxx.colorado.edu/~atlas/math/math1.html> (March 15, 2005).

Charity, Mitchell N. "10^X: An Exponential Notation Meta Page (and scientific notation too)." Vendian Systems. <http://www.vendian.org/mncharity/export1/exponential_notation_meta/> (March 15, 2005).

Feldmeier, John. "Units and Unit Conversion." Case Western Reserve University. <http://burro.astr.cwru.edu/johnf/scales.rev.txt> (March 15, 2005).

Molecular Expressions. "Secret Worlds: The Universe Within." Optics and You (Interactive Java Tutorials). January 17, 2005. <http://micro.magnet.fsu.edu/primer/java/scienceopticsu/powersof10/index.html >(March 15, 2005).

University of Washington. "Scientific Methods: Scientific Notation." Department of Astronomy. February 3, 2000. <http://www.astro.washington.edu/labs/clearinghouse/labs/Scimeth/mr-scnot.html> (March 15, 2005).

Sequences, Sets, and Series

Overview

Sets, series, and sequences are all interrelated. They are each helpful in dealing with groups of numbers, especially large groups of numbers. Sets can contain useful data and can help mathematicians understand these groups of numbers with greater clarity. Sequences always form patterns. These patterns have many practical applications in other areas of mathematics and also in life. From these sequences emerge ratios and formulae that have many applications. One such sequence is the Fibonacci sequence, which is directly related to the golden ratio. This ratio was often used in ancient architecture and is still used today.

Fundamental Mathematical Concepts and Terms

A series is a natural extension of a sequence. However, series can stretch further. The formulae that can emerge from a series have applications in probability and many other areas of mathematics.

SETS

A set is simply a collection of things. These things can be of any genre: numbers, coins, animals, trees, etc. However, in mathematics, a set is usually a set of numbers. The things that make up a set are called elements.

A set contains elements of a specific characteristic. There is usually a reason that elements in a set are part of the set. There is some commonality between the elements. In this way, sets, and sequences are connected.

SEQUENCES

A sequence is an ordered set of mathematical terms. It is usually formed by a specific rule. A sequence can also be called a progression. There are two main types of sequences: arithmetic sequences and geometric sequences.

An arithmetic sequence occurs when the difference between successive terms is the same. For example: 2, 4, 6, 8, 10. The difference between each term and the one before it is 2. Therefore this is an arithmetic sequence. In general terms this can be stated as the nth term minus the $(n-1)^{th}$ term is equal to a constant. The formula for finding the nth term of an arithmetic sequence is: $t_n = a + (n-1)d$.

A geometric sequence occurs when the ratio between successive terms is equal. For example: 3, 9, 27, 81, 243 . . . , where a is the first number in the sequence, d is

Fibonacci Sequence

The Fibonacci sequence is one of the best-known sequences. It was written down, and its properties examined, by Fibonacci in 1202. Fibonacci, also known as Leonardo of Pisa, was an Italian mathematician. The sequence is formed by adding the previous two terms to obtain the next term: $t_n = t_{n-1} + t_{n-2}$. In other words, $t_1 + t_2 = t_3$. The sequence is: 1, 1, 2, 3, 5, 8, 13, 21, 34, 55 . . . *ad infinitum* (continuing forever).

The Fibonacci is directly related to the golden ratio. The ratio between successive terms approaches the golden ratio as the number of terms approaches infinity. The golden ratio is usually written as the Greek letter *phi* (ϕ). The exact value of the golden ratio is

$$\frac{1+\sqrt{5}}{2}$$

The golden ratio is used in ancient architecture. The proportions of the length to the height on the front of ancient Greek and Roman buildings are often the golden ratio. This ratio is proven to be the most aesthetically pleasing ratio. The golden ratio is also found in nature. The nautilus shell spiral can be created by drawing a curve through successive golden rectangles. The pinecone is another example of this same spiral.

the difference between each term in the series and the next, and n goes on as 1, 2, 3, ... In this case the ratio between a term and the term preceding it is 3. In other words the 4th term divided by the 3rd term gives 3. In general, this can be stated as the nth term divided by the (n−1)th term is equal to a constant. The formula for finding the nth term of a geometric sequence is $t_n = ar^{n-1}$, where a is the first term, r is the factor by which each term differs from the one before, and n goes as 1, 2, 3, ...

SERIES

A series is a sequence that is derived from the sum to n terms of another sequence. It can also be defined as the sum of a specific number of terms in a sequence. For the sequence $t_1, t_2, t_3, t_4 \ldots t_n$ the corresponding series would be: $t_1 + t_2 + t_3 + t_4 + \ldots + t_n$. Sequences and series are related, but different.

An arithmetic series is formed from an arithmetic sequence and a geometric series is formed from a geometric sequence.

The sum of an arithmetic series can be found using the formula $\frac{1}{2}n[2a+(n-1)d]$.

The sum of a geometric series can be found using the formula:

$$S_n = \frac{a(1-r^n)}{(1-r)}$$

and the sum to infinity:

$$S_\infty = \frac{a}{1-r}$$

Just as a sequence can be finite or infinite, a series can also be finite or infinite. An infinite series can be convergent or divergent, also just as a sequence.

A Brief History of Discovery and Development

Zeno of Elea, who lived about 490–425 B.C. was the first mathematician to write about infinite series and sequences and their sums. Archimedes discovered a way to show that infinite sequences could have finite results. Chinese mathematicians used methods that have led to an understanding of long-term behavior and limits of infinite sequences. Mathematicians have used sequences over the years to develop new methods in calculus. More recently sequences have found applications in computing.

Real-life Applications

OPERATING ON SETS

A set P consisting of the numbers 1, 2, 4, 8, and 16 is written in set notation as: P = {1, 2, 4, 8, 16}.

To state that 4 is an element of P, the following notation is used: $4 \in P$. $4 \notin P$ means that "4 is not an element of P."

If set A consists of 2 and 4, i.e., A = {2, 4}, then it is a subset of P. This is written in set notation as $A \subseteq P$. Alternatively it can be written as $A \subset P$. However, this means that A is a proper subset of P, which means that it is a subset of P but is not equal to P. To write that A is not a subset of P the following notation is used: $A \not\subseteq P$.

Including all numbers in an infinite set would be impossible. Simple set notation makes this an easy task.

"M is a set containing all integers greater than one" is written as: $M = x \in \text{Integers}(x > 1)$. This reads as "M is a set of all integers x such that x is greater than 1." The vertical bar means "such that." It states a condition. Commas separate multiple conditions.

Occasionally a set will have no elements. This set is called the null, or empty, set. It is represented by the symbol: $\varnothing$. For example: $A = \{\ \} = \varnothing$.

Sets can also be part of a group. That is, data in a set can be part of a larger group of data. There is set notation that deals with multiple sets. A Venn diagram most easily represents these because a Venn diagram is a method of representing multiple sets graphically. $A \cup B$ means "A union B";. This is a way of writing the elements that are in both sets combined. (See Figure 1.) A Venn diagram is a method of representing multiple sets graphically. $A \cap B$ means "A intersection B". This is a way of writing the elements in both set A and set B—in other words, the elements common to both set A and B. (See Figure 2.) These types and/or principles are used in database searches, the most common being an Internet search. A search may be conducted for something containing the word cat AND dog—mathematically this would be cat $\cap$ dog.

Sets are frequently used in everyday applications. Their most common application is in classification schemes, whether it be for clothes, food, animals, or socks. Catalogues are another example of nonmathematical sets. Sets are also used in data analysis in genetics. Chromosomes are sorted and arranged in sets according to shape and specific lengths of chromosome arms and other factors.

USING SEQUENCES

There are finite sequences and infinite sequences. A finite sequence has a limited number of terms. An infinite sequence has an unlimited number of terms. An infinite sequence can be a divergent or convergent sequence. A divergent sequence is truly unlimited. A convergent sequence approaches a limit as the number of terms approaches infinity. Convergent sequences are usually geometric sequences. This is because if the rule for finding the nth term is $t_n; = ar^{1/n}$ then as n approaches infinity the nth term approaches 0. This is because $r^{1/n}; = 1/r^n$.

Sequences always have a specific pattern to them. However, occasionally there is a sequence where a pattern exists that is unknown but it would be beneficial to understand the pattern. An example of this is the stock market. Researchers have been searching for many years to find a pattern behind the stock market, with varying degrees of success.

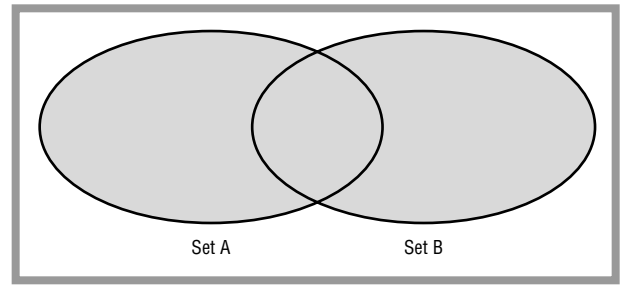


Figure 1: The elements in set A and B.

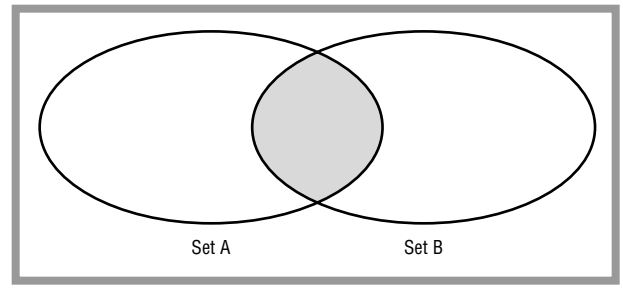


Figure 2: Venn diagram of the intersection of sets A and B (intersection is shaded area).

ORDERING THINGS

Associated with sequences is the notion of a specific order. Placing objects or numbers in an order can give them meaning, such as placing soccer teams on a ladder with the leader at the top and the team with the least number of points at the bottom. This is a simplified sequence.

GENETICS

Sequences are also used commonly in the field of genetics. Specific genes are sequenced (e.g., their base sequence is identified) to determine exactly what gene is associated with a specific physiological function, characteristic, or disease. Sequences of bases determine what gene is formed and what the gene does.

SERIES

A convergent series converges, or comes to, a finite sum. The series $0.5 + 0.25 + 0.125 \ldots$, is a convergent series because even if extended to infinity its sum is finite. Using the above formula it is easy to see that the sum to infinity of this series is 1, and so it is a convergent series. Another way to think of a convergent series is to think of a radio signal that is attenuated by half it's strength each time it pulses (a sequence used in some timing devices). At each step the signal strength decreases but will mathematically never reach zero. Mechanically the signal

reaches a functional zero when the signal strength is too weak to be measured. This is the concept of a sum approaching a specific figure as the number of terms approaches infinity.

Potential Applications

Series can be used to predict regular repeating events, for example, earthquakes and the weather. The data mathematicians collect can be analyzed as a sequence or a series and thus future events predicted in an accurate manner. Sequences can be used in speech recognition. The sound waves produced can be converted to a sequence, similar to a sine wave. Each pitch has its own specific sine wave, which in turn has a specific sequence of y values and x values and these can be converted to a sequence. Sequences also have potential uses in the communications industry, most specifically in signals analysis and wireless connections.

Where to Learn More

Books

Bowerman, Bruce L. *Forecasting and Time Series: An Applied Approach,* 3rd ed. Belmont, CA: Duxbury Press, 1993.

Engquist, Bjorn. *Mathematics Unlimited: 2001 and Beyond.* Berlin/New York: Springer, 2001.

Han, Te Sun. *Mathematics of Information and Coding.* Providence: American Mathematical Society, 2001.

Nelson, David. *The Penguin Dictionary of Matematics,* 2nd ed. England: Penguin, 1998.

Web sites

Drexel University. "Golden Ratio, Fibonacci Sequence." Math Forum. <http://www.mathforum.org> (October 19, 2002).

Rider, J.W. "Mathematical Set Notation." <http://www.jwrider.com/lib/setnotation.htm> (January 21, 2005).

Taylor, S. "Applications of Sequences and series." Algebra Lab: Mainland High School. <http://www.damtp.cam.ac.uk/user/gr/public/gal_milky.htm> (February 10, 2005).

Thomas' Calculus, 10th ed. "History of Series and Sequences" <http://occawlonline.pearsoned.com/bookbind/pubbooks/thomas_awl/chapter1/medialib/custom3/topics/sequences.htm> (February 10, 2005).

United States Naval Academy. "Set Notation." <http://mathweb.mathsci.usna.edu/faculty/traveswn/sm230old/Set%Notation.htm> (January 21, 2005).

University of Cambridge. "Arithmetic Sequence." <http://thesaurus.maths.org> (January 21, 2005).

University of Cambridge. "Arithmetic Series." <http://thesaurus.maths.org> (January 21, 2005).

University of Cambridge. "Geometric Sequence." <http://thesaurus.maths.org> (January 21, 2005).

University of Cambridge. "Series." <http://thesaurus.maths.org> (January 21, 2005).

Weisstein, E.W. "Sequence." Mathworld: Wolfram Resource. <http://mathworld.wolfram.com/Sequence.html> (January 21, 2005).

Sports Math

Overview

Sport, at its best, is a perfect marriage of emotion and execution. The exhilaration of competition and the endorphin-fueled rush of physical fitness are not capable of being measured with mathematical precision. However, the results achieved in sports of every type, whether the endeavor is an individual pursuit or a team game, are invariably assessed in two ways: using the basic math of counting to keep score and assessing the subjective perspectives of the participants about how they regarded the performance.

When the subjective, emotional components of the particular sport are stripped away, the desire to improve performance will often become the focus of an athlete or a team. The desire to improve technique and to achieve better results in competition has spurred the development of numerous mathematical approaches within every sport to measure and to compare aspects of performance.

Math principles are fundamental to sport. They manifest themselves on a number of different levels, from simple counting and keeping track of a score or time, to mathematics as a tool of human discovery as to how a particular sport can be played better.

Sports math can be grouped into three general categories. The first is rules math, in which mathematics is the regulatory basis for the sport. The second grouping is math as an interpretive or demonstration tool in which the understanding and the illustration of aspects of athletic performance is assisted by the application of mathematical concepts, or in which math is used to indicate or predict future performance. The third grouping is performance math, in which mathematics alone, or in conjunction with other science concepts, particularly those of physics, is used to assist in the improvement of athletic performance.

Fundamental Mathematical Concepts and Terms

Counting is fundamental to the appreciation of sport, both in competition and in training. This concept is present in an elemental fashion in sport through the recording of scores, the measuring of distances, and the keeping of accurate time, both as a standard of achievement as well as a competition limit.

Counting in a more sophisticated form is found in most sports through the compilation and use of statistics. Statistics is defined as the branch of the science of mathematics related to the collecting, classification, and use of

numbers, often in large quantities. Statistics in sport are often expressed as percentages.

A percentage is a fraction with a denominator of 100. A percentage may be expressed using the word "percent," as in 25 percent, or using the % symbol, as in 25%. Percentages are the natural mathematical extension of three other familiar concepts: fractions, ratios, and proportions. A fraction is a number written as one whole number divided by another; for example, one half is expressed as ½. A ratio is the relationship between two magnitudes. For example, if the payroll, meaning the total of the salaries paid to all members of a professional baseball team, was $40 million 10 years ago and $80 million today, the relationship between the two payroll figures may be expressed as a ratio of 1 to 2. A proportion is a pair of ratios expressed as a mathematical equation. For example, if a college basketball team has a roster of 15 players, and three of the players are left-handed, the ratio of the team members who are left-handed will be expressed as 3/15, or 1/5.

All percentages are an expression of a relationship based on 100. As is set out in the various sports applications that follow, every fraction, ratio, and proportion may be expressed as a percentage. Percentages may also be expressed where required using decimals, as in the figure 66.92%.

Mathematics is the language of physics. Physics and its particular applications to sport, both in the understanding of the mechanics of human movement as well as in equipment construction, is made clearer in its application through statements expressed as mathematical equations.

A Brief History of Discovery and Development

RULES MATH

How math came to occupy its place as the prime regulatory tool in sport is best understood by the following simple progression: Physical activity, followed by specific activities, followed by informal competition, followed by structured competition, followed by codified rules of competition (time, space, distance), followed by scorekeeping (simple math), followed by performance analysis (advanced math, statistical measures, and mechanics of sport).

At its root, sport is competition, from the informal challenge that individuals make to themselves as they run to keep fit, simply testing themselves personally, to the organized event against a rival or a team of rivals. Competition, to be organized and certain, requires a framework, a structure within which the event can occur with certainty for every participant that everyone is competing in exactly the same fashion. The structure must have limits of time to give the competition a fixed duration, and space or distance to provide boundaries within which the sport can occur.

The evolution of many sports has been accompanied by a progression in the rigidity of rules concerning time and space. Lacrosse, as originally played by Native Americans and referred to as "the little brother of war," was played with teams of hundreds of men, goals set miles apart, in contests lasting as long as three days, and no particular rules of engagement. Timekeeping and precise boundaries were unimportant to the competitors. As another example, ice hockey was originally played on large frozen ponds or rivers, with no lines or markings to regulate play. Modern sport places a greater premium on precise and effective counting of time, space, and distance in the creation of a venue for a competition.

The origins of sports math are best understood in the context of the math applications at the heart of traditional athletics: the track and field competition. Modern track and field events are modeled to a large degree upon the motto of the ancient Olympics of Greece: higher, faster, stronger. Traditional athletics, whether the high jump, the shot put, or foot races ranging from the sprints to the marathon, required the barest of mathematical measures: a defined, accurately determined distance to run, a precisely set object to jump, or an accurately weighed object to throw as well as the measurement of the throw itself.

A timing device to measure performance was a latecomer to sport. By the mid-1800s, the time it took a runner to run a particular distance became important as standards of athletic achievement had become an important public fact. Handheld precision stopwatches were the timing standard in track and field competitions until the 1960s.

The results in these athletic disciplines are calculated in simple mathematical terms, including the fastest time, the furthest distance, the greatest height. Advances in technology have taken the mathematics involved to even greater degrees of precision. For instance, a 100-meter race at the St. Louis Olympics in 1904 took place on a cinder track, where the distance was measured by a steel tape and the races timed by handheld stopwatches. Modern events are run on tracks measured by way of sophisticated electronic means and timing is similarly accurate to thousands of one second. However, the basic mathematics involved in determining the "higher, faster, stronger" concept is not changed.

The geometry of basketball. ELLEN H. WALLOP/CORBIS.

Math as the rules of a sport extends to team competitions. Fundamental to all team games are the following mathematical concepts:

- keeping score
- keeping time
- keeping track of the players
- ensuring conformity with rules concerning size (i.e., a regulation basketball court is 94 feet long; a regulation American football field is 120 yards long; an ice hockey goal is 6 feet wide by 4 feet high)
- keeping a record of the statistics common to the game

Real-life Applications

MATH TO UNDERSTAND SPORTS PERFORMANCE

Beyond the result achieved in any competition, numbers are used throughout sport to explain and to better understand performance. Statistics are widely relied upon in virtually every team sport as a means of enhancing the understanding of both individual competitions as well as entire seasons, for both the participants and the public at large. The interpretation of sports performance through mathematics will often provide an understanding of a result that the scoreboard does not reflect.

Statistics in sport must be regarded with caution. Media commentators often speak with apparent authority about a sport through their reliance on the numbers that are associated with competition. An understanding of the mathematics involved in these interpretive aspects of sports math is critical to separating the statistical "wheat from the chaff."

There are levels of interpretation that mathematics can provide. Some simple statistics provide a peephole on performance, others a picture window. The math involved is not the entire story, but merely an insight into the actual result achieved. The better the relationship between the math and the subject sport, the more likely the analysis is more insightful.

The following statistics are examples of how sport observers, especially in the media, often present

arguments to support a contention that a certain player is outstanding:

- An NBA basketball player averages 22.0 points per game over the course of a season. Is such a player an elite performer?
- A major league baseball pitcher wins 18 games in a season, and he has an earned run average calculated at 4.5 runs per game. Is this pitcher an elite player?
- An NHL hockey goaltender has a save percentage of 0.93 (for every 100 shots on his goal, he saves 93). Is this athlete an elite level player?

In each of these three examples, the answer to the question of whether each player is an elite must be a resounding "maybe" or "perhaps." The NBA player scoring 22.0 points per game may be a terrific all-round player, with his strong scoring average one facet of his skills, or he may be a weak defensive player who is a liability at his end of the floor. The pitcher might be truly dominant, or he may be one of those athletes with impressive overall statistics, who performs poorly in important situations in critical games. The NHL goaltender could be a steady performer, or might be inconsistent, getting a shut-out one game and allowing six goals the next. Without a more in-depth and focused use of mathematical principles, these simple numbers are not the basis for great insight when taken by themselves.

In most team sports, such as baseball or basketball, individual statistics are an indicator, and not a determinant, of team success. The final score in a team contest is the only absolute measure of the team's success on a given day.

Sport statistics are almost universally expressed as either a percentage, or as a decimal. In sports, the score in a game or the result achieved by an athlete in an individual competition represents "what." Various kinds of mathematical and statistical applications can tell a great deal about the related questions of "why" and "how." Math is variously an interpreter of past performance and a predictor of future performance. The game of baseball provides some useful insights in this regard.

BASEBALL

Baseball is a sport saturated in statistics. Its fans will often exaggerate when explaining the nuances of baseball to more casual observers that this sport is the only major team sport played without a clock, which is said to give it a subtlety that can be captured by a wide variety of statistical measures. Further, because baseball has inherent repetitions of a number of actions within the game over a season (a major league player may face a pitcher more than 500 times in a season, make hundreds of throws, and run the bases hundreds of times), each player and each team perform in ways that can be readily converted into a statistical measure.

Some baseball statistics, such as batting average and earned-run average, are ingrained in the public consciousness as key indicators of performance. Over the past 30 years, the desire to delve further into the interpretation of baseball performance led a number of statisticians to develop the analysis into a field now known as sabermetrics. This analysis extends the mathematics of baseball from the relatively simple formulae of batting average and earned-run average to detailed calculations used to both rank player ability and to predict future performance.

Baseball batting averages are expressed as percentages. A player's batting average is the number of base hits made by the player divided by the number of at bats (an "at bat" in baseball is defined by the number of times the player comes to bat, less all walks, errors, and times hit by pitch ball). For example, Player A has 140 hits, 30 walks, was hit three times by a pitch, and reached first base five times as a result of errors in a season. He took 220 total bases. He has 600 at bats. His batting average is 140 / 600 = 0.233. In baseball terms, Player A is said to be a "233 hitter."

The on-base percentage statistic determines how effective a player is at getting on base, by all possible means. The on-base percentage is Number of times to first base / Number of plate appearances = 140 + 30 + 3 + 5 / 600 + 30 + 3 + 5 = 178 / 638 = 0.279. One would conclude from this calculation that Player A's simple batting average is deceiving; Player A is more effective, by the ratio of 0.279 to 0.233, at getting on base by any means than the simple batting average statistic reveals. As the object of baseball is to score more runs than one's opponent, getting on base, and being in a position to score a run, is a more accurate statistical measure of a player's worth to a team than the number of hits the player may collect.

For many years, baseball placed a premium on the total runs scored by a player as a measure of effectiveness. Sabermetrics took this analysis in a more detailed direction, that of "runs created" by an individual player. As in the on-base percentage example, the object of baseball is to score more runs than the opponent. If determining how effective a player is at getting on base is a more useful indicator of a player's contributions than simple batting average, determining how effective a player is at creating runs will be an even more useful statistic to analyze the abilities of a given player. The player who can create the most runs, in whatever fashion, will likely be the most valuable offensive player on a team.

Total bases are the number of bases that the batter's hits amounted to over the season, with a single being one

base, a double two bases, a triple three bases and a home run all four bases. In an individual game, for example, where a player hit two doubles and a home run, the player would have eight total bases. Therefore, Runs Created is (Hits + Walks) × Total Bases / (At Bats + Walks). With this equation, Player A's statistics read (140 + 30) × 220 / (600 + 30) = 37,400 / 630 = 59.37.

Using this analysis, Player A created approximately 59 runs for his team over the season. If the team scored 700 runs, and if 12 players had been used regularly as hitters in the team's line up over the course of the season, Player A is proven to be a slightly above average contributor to his team's offensive production (700 team runs divided by 12 players means that the average regular hitter would be expected to create 58.3 runs). To complete the progression from the simple batting average calculation to the runs created determination, a relatively modest .233 batting average translates in the example to a slightly better than average contributor to the offence of this team.

As will be further illustrated, and as the above baseball examples confirm, the rough rule of thumb regarding sports statistics is that the more intricate and involved the desired analysis of athletic performance, the more typically involved the mathematics will be.

NORTH AMERICAN FOOTBALL

Football is a game of territorial conquest, and the measurement of the amount of territory gained by competing teams has been a focus of performance analysis in this sport. Like the evolution in the extent and the sophistication of mathematical applications in baseball, North American football has grown from a game where bare statistical measures of leading scorers and yards gained by individual players have given way to involved analyses of every aspect of the game. Thirty years ago, a quarterback, the single most important player on a team, would generally be assessed on the following set of statistical indicators:

- Completion percentage: The number of passes completed divided by the total number of passes attempted; percentages in the range of 55–60% were typically considered good.
- Number of touchdowns versus the number of interceptions: A good quarterback typically would throw for more touchdowns than interceptions.
- Yards gained through total number of passes: In a territorial game, quarterbacks who passed for a greater number of yards were generally more valuable.

With these statistics, it was possible to have an understanding of the performance of an individual quarterback, but the statistics did not give a complete picture. For example, a quarterback on a team that threw the football less frequently might appear to be an inferior player when compared with a player whose team threw the ball a great deal. Using years of data, analysts were able to incorporate known and reliable statistics to create useful mathematical tools with which to assess performance, as well as to establish a standard that would permit comparison between individual players.

The desire to better understand quarterback performance lead football analysts to develop the Quarterback Rating Index, which is calculated as follows: (1) Total pass completions divided by total pass attempts; (2) Subtract 0.3 from (1); (3) Divide by 0.2 and record the total (the result cannot exceed 2.375 or it may be less than 0), and this gives Subtotal one; (4) Total passing yards divide by total pass attempts; (5) Subtract 3 from (4); (6) Divide by 4 and record the total (the result cannot exceed 2.375 or may it be less than 0), and this gives Subtotal two; (7) Total number of touchdown passes divided by total pass attempts; (8) Divide result by 0.05 and record the total (the result cannot exceed 2.375 or it may be less than 0) and this gives Subtotal three; (9) Total number of interceptions divided by total pass attempts; (10) Subtract 0.095 from (9); (11) Divide the result in (10) by 0.04 and record the total (the result cannot exceed 2.375 or it may be less than 0), and this gives Subtotal four; (12) Add the four subtotals recorded; (13) Multiply by 100; (14) Divide by 6. This final total is the Quarterback Rating.

This more complicated Quarterback Rating Index is seen as more reliable because it incorporates every aspect of the quarterback's ability to pass into the equation. It takes the analysis beyond interpretation into an understanding of the individual player's performance.

However, as with any statistic that is not the final score or result, even the complicated ratings of this Index are not a compete picture. If a quarterback throws for three touchdowns and many yards after his team is hopelessly behind, the individual rating may be enhanced, but the team performance not assisted. An interception thrown in the first quarter of the game is given the same weight as an interception thrown in the last minute of a tied game that is returned by the other team for a winning score. As with all forms of statistical analysis, the factors not calculated in the statistical equation must be assessed as well.

BASKETBALL

Basketball is a simpler, more free flowing game than either baseball or North American football, and its statistical base has long been the individual statistics of players, totaled for team assessment. In basketball, statistics are

seen as a measure of tendency, as opposed to the interpretation of individual performance. For all of the importance attached to scoring averages of players, basketball experts key on two chief statistics: rebounding (broken into offense and defense) and free-throw shooting.

Studies have illustrated that where a team has more rebounds than an opponent, that team will be expected to win 70% of its games. When that same team is more effective from the free-throw line, the success rate for the team is between 85% and 90%. These statistics bear out the nature of the game itself, i.e., rebounding advantages mean that a team is controlling the ball on defense and likely getting more than one shot on any given sequence on offence. Good free-throw shooting means that the team is getting in a position on offence to take shots, drawing fouls from the opponent, an indication that the team is better controlling the ball and the game than its opponent.

These statistics also underscore the fact that each sport has subtleties that are inherent in the interpretation of the statistics gathered, and each sport has its own measure of what constitutes success. In baseball, a professional hitter who has a batting average of .350 is very likely a successful offensive player; an NBA basketball player with a free-throw shooting percentage of 35% would be considered a very poor free-throw shooter; an NFL quarterback with a passing completion rate of 35% would not succeed as a player at that level.

PREDICTING THE FUTURE: CALLING THE COIN TOSS

In many respects, the coin toss is a metaphor for life itself: if one's call is wrong on one occasion, one will likely get another chance at some later time. The coin toss in North American football is one of the great rituals of the game; the winner of the coin toss has the right to elect to receive the opening kickoff and thus take initial possession of the ball. The loser of the coin toss can select which end of the field they will defend. In a game that is, at its core, one of territorial conquest, success at the opening coin toss in a football game is important, especially given the variables of wind, field condition, and weather where the game is being played outdoors. The direction and the outcome of the game may well be influenced by the result of the coin toss.

In the American professional game, the coin toss takes on special significance if the game is tied at the end of regulation time. To determine which team will take possession of the ball at the commencement of sudden death overtime, where the first team to score in any fashion wins, the referee will toss the coin and the winner of the coin toss will inevitably elect to receive the kickoff.

The question is, Is there a mathematical predictor for how the coin toss should be played by either team? The probability of a coin being heads or tails is 1:1, or an even chance, every time. No matter that, for example, the previous five coin tosses may have been heads, on the sixth and subsequent tosses the probability of heads versus tails remains even. In other words, each coin toss must be approached as a unique, free-standing event: any history will be irrelevant.

The probability of heads (or tails) is expressed as Number of Favorable Outcomes / Number of Possible Outcomes = ½. The probability of two heads (or tails) being tossed in a row is $\frac{1}{2} \times \frac{1}{2} = \frac{1}{4}$. The probability of four heads (or tails) being tossed in a row is $\frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} = 1/16$.

PASCAL'S TRIANGLE AND PREDICTING A COIN TOSS

The French mathematician Blaise Pascal (1623–1662) is regarded as the developer of the device known as Pascal's triangle (although history confirms that Chinese mathematicians developed a very similar construct 500 years before Pascal). Pascal's triangle has a number of algebraic applications. (See Figure 1.)

To read this table in terms of coin toss probabilities, suppose that the number of tosses is one. On the corresponding line, the triangle provides the numerators for two possible outcomes, heads or tails. As the total number of outcomes is two, the denominator for this calculation shall always be two, meaning that the probability of heads or tails is ½.

When the coin is flipped twice, there are three possible outcomes: (1) heads once, tails once, (2) twice heads, and (3) twice tails. The extreme possibilities, two heads or two tails, are represented by a "1" on each side of the triangle. The one head, one tail result (which may occur in two different orders), is represented by "2." The extreme possibilities are therefore ¼, and the probability of one head, one tail is 2/4, or the expected 50%.

If the third line of the Pascal's triangle in Figure 1 is examined, the total number of possibilities is $2 \times 2 \times 2 \times 2$, or 16. By using the triangle as a calculator, the probabilities can be determined as All heads, 1/16; One head, Three tails, 4/16; Two heads, Two tails, 6/16; Three heads, One tail, 4/16; All tails, 1/16. (As the triangle is symmetrical, it does not matter which side is called heads or tails.)

There has been interesting research carried out recently that suggests that the side of the coin facing up in a coin toss is slightly more likely to be the side turned up on the flip. The premise of such research appears to be

that as coins are tossed by real people, who are subject to bias, and the person tossing the coin sees the head or tail facing upwards, they may tend to unconsciously catch the coin with that face up. The probability is said to rise to 0.51 in favor of the exposed side as opposed to the accepted 0.50. This theory will no doubt be the subject of further study.

FOOTBALL TACTICS—MATH AS A DECISION-MAKING TOOL

In many team games, math is a tool for strategy decisions, many of which are rooted in the concepts of probability. The basic question that is often addressed is, in a given situation, what play or strategy affords the best chance of success?

American football is a game of field position and territorial advantage. As a general rule, the team with the consistent best position on the field will be in the best position to score and therefore win. A common tactical decision in American football related to field position is, given where the team has the ball on the field, whether the should team punt the ball away to the other team, attempt to gain a first down and keep possession of the ball and ultimately score, or kick a field goal.

Based on the data gathered from more than 700 NFL football games, the following statistical analysis can be made: Team A has a fourth down on the Team B 2-yard line. Team A is assessing its options: punt, attempt to score a touchdown, or attempt a field goal. A punt by Team A is not a sensible option, as the ball would be kicked through the 10-yard end zone and would be placed at the Team B 20-yard line, where Team B would take over on offense. The two realistic options for Team A are the attempt to score a touchdown or to kick a field goal.

At the NFL level, the statistical data confirms that an attempt to score a touchdown from the 2-yard line, coupled with a virtually automatic extra-point conversion, has a probability of success of 40%. Therefore, the value of that choice can be calculated as 6-point touchdown $\times$ 0.40 = 2.4 points; 1-point convert $\times$ 0.40 = 0.4 points. Therefore, the total value of attempt = 2.8 points.

An attempt at a 3-point field goal from the 2-yard line is as likely to be successful as the 1-point conversion after a touchdown (the success rate is slightly under 99%). The value of the field goal choice is 3-point field goal $\times$ 0.99 = 2.97, which is in essence 3 points.

Using the value of each choice with the probabilities for each calculated, it would seem that Team A would have a slightly better option with the field goal over the touchdown attempt (3 points versus 2.8 points). However, as stated above, American football is a game that turns, to a large degree, on field position; the decision as to attempt a field goal versus the try for a touchdown must also be assessed considering the field position consequences that flow from each choice.

Number of Coin Tosses	Numerator for Probabilities	Denominator for Probabilities
1	1 1	2
2	1 2 1	4
3	1 3 3 1	8
4	1 4 6 4 1	16
5	1 5 10 10 5 1	32
6	1 6 15 20 15 6 1	64
7	1 7 21 35 35 21 7 1	128
8	1 8 28 56 70 56 28 8 1	256

Figure 1: Pascal's triangle as a probability tool.

By rule, after a successful field goal, Team A would kick off to Team B. On average, statistics confirm that an NFL kickoff will be returned to the receiving team's 27-yard line. A first down and 10 yards to go situation at that position for Team B is worth 0.6 points to Team B, based on the probabilities of scoring from that position.

If Team A were to attempt to score a touchdown from the Team B 2-yard line and fail, the ball would be turned over to Team B at the same 2-yard line. With Team B 98 yards from the Team A's goal line, this poor field position statistically is worth −1.6 points to Team B. With the field position components factored into the calculation weighing the attempt at a touchdown versus a field goal, the value of each choice can be calculated as Value of Touchdown Attempt = 2.8 points; Field position (−1.6 points to Team B) = 1.6 points to Team A; Total Touchdown Attempt Value = 4.4 points; Value of Field Goal Attempt = 3.0 points; Field Position (+0.6 points to Team B) = −0.6 points to Team A; Total Field Goal Attempt Value = 2.4 points.

With the field position information now factored in, it is apparent that what was a slightly preferable course of action, the field goal attempt, is now a significantly lesser option when compared to the touchdown attempt.

As with any mathematical model employed to predict an event or to select the optimum course of action, there will be variables that cannot be reduced to a number or an equation. In the above example, factors, such as how much time is left in the game, field conditions, or an

injury to a key player on either side, would potentially alter the decision that the probability calculation otherwise directs as the best choice. Math is rarely determinative with respect to decisions in team sports, but it is often very illuminating.

UNDERSTANDING THE SPORTS MEDIA EXPERT

It is standard in the television coverage of virtually every sport, from professional competitions to the various events in the Olympics, for the broadcasting network to engage the services of an expert to assist in the presentation of the event. A typical broadcast will have a commentator giving the audience a play-by-play of the action being telecast, and the expert, often referred to as a color commentator, provides insight and analysis about the event, both as it is unfolding and at various stoppages in play.

In addition to what the expert may be saying to the audience, it is common for the presentation to display statistical summaries in relation to either the individual players, the teams, or the season to date. Where the casual fan is seeking information to enhance their enjoyment of the broadcast and the game itself, the numbers cited by the experts are often not helpful, but can actually be confusing. It is important to approach these statistics-filled commentaries with caution.

For example, in professional baseball, basketball, and hockey, it is common for a playoff series to be played as a best-of-seven-games event. In a baseball World Series, where one team is ahead of its rival three games to none, there will inevitably be an expert commentator who might suggest that "never in World Series history has a team come back from three games down to take the Series." This statement might be true, a powerful sounding pronouncement, which can be examined more closely using math principles to test its weight.

Between 1920 and 2004, there have been only 20 World Series where a team led 3-0. Assuming for a moment that all other factors are equal, and that the two teams are relatively evenly matched, the odds of a team winning four straight games can be calculated using the same probability as the coin toss analysis or Pascal's Triangle ($\frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} = 1/16$). Based on this calculation, one would therefore expect a four-game comeback to take a series as very rare. Mathematically, one would not expect this result with high probability in 20 World Series. Again, variables such as talent disparity, injury, the tendencies of teams in certain stadiums, and similar factors will play their role. The basic math, however, underscores that the expert's breathless pronouncement about the difficulty of a comeback is an overstatement.

Another common example of a statistic used in a superficial fashion is the emphasis placed by the expert on an aspect of the game that is not essential to a team's success, yet stated in an authoritative fashion. It is common in the course of an NBA basketball season to hear references to a certain player being the best slam dunker on his team or in the league as a whole. A dunk, or slam dunk, is the delivery of the basketball by a shooter through the cylinder to score with the ball being propelled down after the player has jumped high enough for the shooting hand and ball to be above the rim.

For example, a statement such as "X is the most prolific dunker in the NBA" might be made by the expert analyst. While there is no question that the dunk is in many situations an emphatic and athletic maneuver, from a mathematical perspective, assessing the relevance of this statistic to team success, this expert statement is of little value, because the dunk is worth the same as any other 2-point field goal attempt in basketball, therefore the manner in which the basket is made has no greater effect on the scoreboard. Further, the dunk is less important than the 3-point shot (a 50% greater value per successful attempt). In a team game, the analysis as to how the ball got to X for the dunk is more important than the dunk itself—how did X become open to make the dunk, what passes or other maneuvers were made by X or his teammates. In basketball, rebounding the other team's miss is a typical way that the ball begins its path to the other team's goal. Therefore, one would expect rebounding to be far more important in an analysis than is a dunking statistic.

In the simple analysis above, one would conclude that rebounding is far more important than dunking. In fact, based on statistics gathered using NCAA men's and women's basketball data over a five-year period, the simple analysis is confirmed. Where a team out-rebounds its opponent, it will win 72% of its games. Where that team also shoots more free throws than its opponent, its success rate climbs to almost 90%. Dunking, dramatic as it may be, is not a significant factor in team success.

Every sport has its statistics that, when employed in commentary, may impress but not necessarily inform the audience. In NHL hockey, frequent references will be made during telecasts as to how hard a particular player can shoot the puck. There is no question that from a physical standpoint, it is an impressive feat for a player to be able to deliver a shot towards the opposing goal at speeds in excess of 100 miles per hour. However, much like the dunking example in basketball, this statistic does not really contribute to the understanding of the game, especially if the information is taken in isolation.

The object of the game is to put the puck in the opposition goal, it is shooting accuracy that will be at a premium.

RATINGS PERCENTAGE INDEX (RPI)

Determining a winner in a team sport competition on one level is an easy task, i.e., the score at the end of the game is the sole indicator of success. In a series or in a season of competition, the winner is readily determined by the season standings, and so long as all teams in the competition have competed the same number of times against all others in the league, the final standing will be the determinate as to the champion.

Standings, and what are referred to as win/loss records, are a less useful standard where there are multiple teams and all do not compete against all of the others in a given season. For example, in American college sports, there may be as many as 350 teams in a particular division. While each team may play in a conference of between eight and 15 other programs, assessing a team using its won/loss record alone among conference rivals is straightforward, but to compare the team regionally or nationally from its conference play is problematic. For example, if a basketball team won 27 games and lost one, as opposed to a team they did not play who compiled a record of 16 wins and 12 losses, the team with the best record is not necessarily the better team, if its opponents were weaker than those of the 16 win team.

When an issue such as the determination of a national champion in a particular sport is at stake, consideration is given as to how teams that did not compete against one another in the regular season might be compared and ranked to create a fair championship that included all deserving teams. The Ratings Percentage Index (RPI) was created to achieve this objective.

The RPI is calculated in different ways for different sports, so as to reflect the nuances of that particular competition, but the RPI is an algorithm that takes into account the common features of a team's season record, the record of its opponents, and the record of its opponents' opponents. The theory in the construction of the RPI is simple: wins and losses must be assessed on a qualitative basis as much as they are counted on a quantitative basis. To calculate the RPI, Team A RPI = 25% (Team A record) + 50% (Team A's opponent's record) + 25% (Team A's opponent's record).

There are a number of variables that the RPI does not address. The following examples show the RPI as a potential rectifier of disparity that appears from a review of comparison teams' win/loss records alone.

Team A and Team B are both NCAA Division 1 women's basketball teams. Each is being considered for a place in the elite National Championship tournament. Team A is from California and Team B is from Connecticut, and the teams play in different conferences. Team A and Team B did not play one another during the course of the regular season.

Team A had a very successful season, compiling a record of 26 wins and five losses, for a wining percentage of 83.8%. Team B struggled for large parts of the season, achieving a record of 16 wins and 14 losses, 53.3%. Team A was therefore more than 30% more successful at winning games than Team B.

For example, Team A winning percentage (83.8%); Opponent's winning percentage (45%); and Opponents/Opponents (58%), the calculation is RPI Team A = $(83.8 \times 0.25) + (45.0 \times 0.50) + (58.0 \text{ x } 0.25) = 20.95 + 22.50 + 14.50 = 57.95$ (round to 58.0).

Continuing the example, Team B winning percentage (53.3%); Opponent's winning percentage (75%); and Opponents/Opponents (60%), the calculation is RPI Team B = $(53.3 \times 0.25) + (75.0 \times 0.25) + (60.0 \times 0.25) = 13.33 + 37.50 + 15.00 = 65.83$ (round to 66.0).

It is evident from this analysis that while Team A had a far more successful season in terms of winning games, Team B played a much more difficult schedule. The RPI would lead to the conclusion that if one were to choose between these teams, Team B is likely the better team. However, while the RPI is a more involved calculation than simple wins and losses, it has significant variables that cannot be reduced to mathematical equation. Those variables include injuries to key players, whether the wins were early or later in the season, or the margin of victory (the RPI calculation treats as equal a win by 25 points and a win in overtime by a single point, the margin is not relevant to the RPI).

As the RPI is used as a statistical measure more frequently, individual sports can customize the calculation to reflect a feature in its game. For example, in 2005, the NCAA adjusted its RPI for basketball championship calculations. The NCAA concluded that as a home court was a significant advantage to the home team in college basketball, a visiting team deserved extra credit for a victory in a hostile environment. The RPI was thus adjusted so that road win = 1.4 wins; road loss = 0.6 losses; home win = 0.6 wins; home loss = 1.4 losses; neutral site = 1.0.

NCAA hockey adopted a similar approach to fine-tune its RPI, with a road win worth 1.5 wins, a neutral site win valued at 1.3 wins, and a home win worth 1.0.

The RPI will never be determinative of ability; arguably, the only reliable measure of that standard in team sports is head-to-head competition. Used in conjunction

with other tools, the RPI provides insight to complicated ranking and seeding issues.

MATHEMATICS AND THE JUDGING OF SPORTS

In most athletic disciplines, mathematics will illuminate aspects of performance. As noted in the discussion of team sport statistics, the more involved the mathematics, often the greater degree of insight into present and future performance.

Mathematics is a rather poor tool when used to explain sports that are subjectively assessed, such as figure skating, diving, synchronized swimming, and other similar disciplines. However, it is important to understand what is being sought to be achieved in the scoring of these disciplines, and to be careful in attempting to interpret any result beyond the obvious ranking of the participants.

Figure skating has always posed particular difficulty for judges: How can a subjective opinion, however knowledgeable, concerning elements of beauty, presentation, and grace be reduced to a score, a hard, certain mathematical proposition? Figure skating has had a number of judging scandals, usually turning on improper collaboration among judges to guarantee that certain participants would achieve certain scores. In 2004 the international figure skating adopted a scoring system referred to as a "code of points." The general proposition of this system is that every aspect of each skater's performance, every jump, spin, turn, and movement will be graded individually. The judges, typically numbering eight at an international event, would then add or subtract points for the skater's execution of each part.

This grade of execution applies to five overall components; the maximum score available in each component is 10.0. Each judge submits their individual score for each skater. The highest and the lowest scores from the judging panel are discarded, with the maximum score attainable of 50.0.

Similar to the manner in which team sports telecasts communicate statistics that are purportedly used to describe performance, but without proper reference to the fundamentals of the sport itself, figure skating judging and the mathematical results generated are difficult to understand. The mathematics here is an imperfect attempt to put a hard number to a purely subjective discipline.

Unless one has an in-depth knowledge of the judging criterion, the viewer is left with a number that is disconnected from performance. If a swimmer races 100 meters in a pool in 55 seconds, that result is observable. If a football player scores a touchdown by running with the ball, that result is observable. The math underlying the scoring in the judged sports is only an indicator to the most expert in that discipline; the more casual fan must treat the numbers generated as a simple ranking.

MATH AND THE SCIENCE OF SPORT

The aim of mathematical applications in sport is not to change the sport in question, but to better understand it. One must be able to understand the essence of performance. As already discussed, mathematics is a very helpful interpretive tool in sport. Math as the language of science can be applied in virtually every sport to understand how the game is played. However, the following examples of math explain how a particular sporting activity is performed, or how it might be performed better.

BASEBALL

The home run is arguably the most dramatic play in baseball, the product of a one-on-one confrontation between hitter and batter. Science will assist in the understanding of a number of features concerning how far a baseball can be hit. Assuming that the bat is constructed of wood and is 32 ounces in weight, that a ball is pitched at 85 miles per hour (an approximate average speed for a pitch thrown by an American major league pitcher), and that the ball was struck squarely on the "sweet spot" (the optimum part of the bat for striking the ball, given that the end of the bat is moving more quickly than the handle). To send the ball 400 feet, the bat must strike the ball at a speed of 70 miles per hour.

To take the analysis one stage further, the difference between the properties of an aluminum bat and a wooden bat can be examined. This analysis will turn on a calculation known as determining the coefficient of restitution (COR), a determination of how "springy" the surface of each bat is, which will impact upon how much energy will be lost in the transfer from the pitched ball to the bat when it is struck.

Due to the nature of each bat's construction, a 32-ounce aluminum bat will have a barrel of 2.75 inches (the maximum size permitted by major league baseball); the wooden bat will have 2.50 inches, as it is not possible to have the weight distributed more to the barrel of the bat, with the thin handle, as the bat tends to shatter on impact with a ball.

The aluminum bat has a barrel circumference that is 1.21 times bigger than the wooden bat (the ratio of 23.74 to 19.63), which translates into a surface available to strike a ball that is approximately 10% larger than the wooden bat. Precise scientific trials using standard baseballs show that 25% of the energy created in the collision

between an aluminum bat and ball will be restored to the ball, as opposed to 20% of the similar energy generated between a wooden bat and a ball. The COR for the wooden bat is 0.45, the COR for the aluminum bat is 0.50. Using this general relationship (and assuming that other variables such as elevation, wind speed, and direction and the angle at which the bat strikes the ball are not factored), if a ball were hit 380 feet with a wooden bat, one would expect an aluminum bat to generate a hit of approximately 410 feet.

It is this type of analysis that has persuaded the authorities to forbid the use of aluminum bats in North American major league competition. The physics of the aluminum bat would not only make home runs an easier proposition, the speed of the ball created on impact with the aluminum bat would create a greater risk of danger for those fielders closest to the batter, including the pitcher, the first baseman, and the third baseman. As noted, the coefficient of restitution is a measure of the springiness of a surface, expressed as the ratio of the speed of the object before and after collision. If a rubber ball were thrown against a wall at 50 miles per hour, and it returned at 30 miles per hour, COR = 30/50 = 0.6.

The following will illustrate why the pitcher and his first and third base teammates are at greater risk from a ball hit with an aluminum bat. The regulated distance from the batter to the pitcher is 60 feet, 6 inches. However, on the delivery of a pitch, the pitcher must maintain contact with the point of measurement (the pitching rubber) only as the ball is being delivered, and as the pitcher throws the ball, the natural pitching motion will carry the pitcher one stride closer to the batter, plus any extra distance that is created by his follow-through. It is safe to assume that the pitcher will be no further than 54 feet from the batter as the ball is released.

Pitches at the major league level vary in speed, but a typical pitch will travel approximately 85 miles per hour (assuming a constant speed, although in practice the pitch will be faster on delivery and slower as it has traveled to the batter's box). The pitch thrown at 85 miles per hour is traveling at a speed of 124.67 feet per second. To travel the 60.5 foot distance to home plate, the ball will reach home plate in approximately 0.49 seconds. Studies using wooden bats confirm that the approximate time for a hard-line drive delivered directly at a major league pitcher is approximately 0.40 seconds.

Studies at the American college level, where aluminum bats are legal for use (with players presumably not as strong or as skilled as major league players), confirm the following: At speeds off the bat at between 103 and 113 miles per hour, batted balls were reaching the pitcher 54 feet away at 0.357–0.315 seconds. It would be reasonable to conclude that a major league batter would deliver the ball harder and therefore in a shorter time than the college sample.

As it is generally accepted that human reaction time is rarely faster than 0.20 seconds to an event, even where the event is anticipated (as with a pitcher who might expect a ball to be struck at him), not only would balls travel further when hit by the aluminum bat, the difference in the available reaction time to the pitcher and the time for the ball to reach the pitcher from the bat when struck would decrease to as little as 0.10 seconds in major league play.

CYCLING—GEAR RATIOS AND HOW THEY WORK

Cycling, at both the recreational level as well as international competition level, requires an understanding of a number of physical principles. The gears used by cyclists are an example. The "penny farthing" bicycles of the period from 1870 to approximately 1900 were constructed with a huge front wheel and a tiny rear wheel, with the pedals connected to the front wheel only. The penny farthing did not have gears connecting the large front wheel and the small rear wheel, and it depended upon the fact that one pedal by the rider would create one rotation of the large front wheel. The penny farthing construction is identical in principal to that of a child's tricycle. As an example, if a tricycle front wheel, to which the pedals are directly connected, is 16 inches in diameter, that wheel will have a circumference 50 inches. One revolution of the front wheel means that the tricycle travels 50 inches. If the child pedaling the tricycle pedals at a speed of 60 revolutions per minute, or one revolution per second, the tricycle will be traveling at 50 inches per second, which is a speed of 2.8 miles per hour. Greater speed can only be achieved through greater revolutions per minute.

If an adult wished to ride a tricycle at a speed of 15 miles per hour, a typical speed at which to ride a bicycle in a recreational fashion, the tricycle front wheel would have to be very large or the cyclist would not be able to generate enough revolutions per minute to get distance. If the adult pedaled at 60 revolutions per minute, to achieve a speed of 15 miles per hour, the front wheel would necessarily be 84 inches, or 7 feet, in diameter. Gearing was necessary to make cycling a more efficient form of movement.

The concept of gearing for a bicycle was first theorized by Leonardo da Vinci (1452–1519) in the 1500s, but the concept was not developed for commercial application until Frenchman Paul de Vivie, alias "Velocio," (1853–1930) built the first functional derailleur, the device that permits the gears on a multi-speed bicycle to be changed

by the rider as the bicycle is ridden. The gearing on a bicycle permits a cyclist to use energy more efficiently to climb hills, and to maximize speed on a downhill.

A typical bicycle has wheels that are 26 inches in diameter. Gears are typically measured by the number of teeth each gear has on its circumference. For example, if a front gear ring on a bicycle has 54 teeth, and a rear gear wheel has 27 teeth, every time the front wheel rotates, the rear wheel will rotate twice, creating a 2:1 gear ratio.

The lowest gear ratio on a bicycle might be a front chain wheel with 22 teeth and a rear chain wheel with 30 teeth. This creates a lower gear ratio of 0.73:1. For each pedal stroke, the rear wheel will turn 0.73 times, meaning that the bicycle will move forward approximately 60 inches (approximately 3.4 miles per hour if the bicycle is pedaled at a 60 revolutions per minute rate). The highest gear ratio on the bicycle might be a front chain wheel with 44 teeth and a rear chain wheel with 11 teeth, creating a 4:1 gear ratio. As the bicycle wheels are 26 inches in diameter, the bicycle will move forward 326 inches with each pedal stroke. If the cyclist pedaled at a rate of 60 revolutions per minute, the bicycle will be traveling at a speed of 18.5 miles per hour. If the pedaling rate were doubled to 120 revolutions per minute, the bicycle would travel at a speed of 37 miles per hour. Gearing permits the bicycle in this example to travel at speeds ranging from 3.4–37 miles per hour, to climb steep hills or to travel along level terrain.

The gearing of the bicycle must be considered along with the cadence or the rate at which the bicycle is pedaled. A smooth, even cadence means that the bicycle will be pedaled efficiently, as the energy expended by a cyclist is delivered most efficiently at certain cadences. For example, on the very steep terrain where mountain bikes are effective, a cyclist may climb a hill at as little as 50 revolutions per minute, whereas a road racer will typically operate at cadences of between 80 and 120 revolutions per minute.

SOCCER—FREE KICKS AND THE TRAJECTORY OF THE BALL

The free kick, bent with skill around a wall of defenders, is one of the most dramatic aspects of soccer. The success of this tactic is dependent upon a host of variables, including the distance from the goal, the height of the opposing players' wall, the height of the goal, the force of the kick, the spin imparted to the ball, the air flow around the ball in flight, and lesser variables, such as air temperature, humidity, and the friction of the grass when the ball is struck.

Using a famous free kick goal from the England vs. Greece 2001 World Cup qualification game as an example, English midfielder David Beckham scored a goal analyzed as: The ball was kicked at 36 m/sec. (80 miles per hour); the ball was kicked from a distance 27 meters from the goal; the ball moved laterally in flight 3 meters (from Beckham's right to left, facing the goal); the ball cleared the defender's wall by 0.5 meters; as it entered the goal, the ball was traveling at a speed of 19 m/sec (42 miles per hour).

The analysis assists in understanding how players taking a free kick can best strike the ball for a similar effect. The high speed of the initial kick results in the ball having very little drag as it moves through the air. However, as it passes the opposition wall, the ball begins to slow, entering a smooth airflow (laminar) phase of its travel. Greater drag on the ball now occurs, when coupled with forces generated by the spin imparted on the ball by the foot of the player (Magnus force), the ball will appear to bend and dip, fooling the goalkeeper.

There have also been various studies conducted in recent years in world soccer due to the rise in the importance of the penalty shot. The penalty shot awarded for a foul committed against an offensive player in the defender's penalty area has long been a feature of soccer. Only since 1982 has the penalty shootout been the sanctioned method of deciding an international soccer game. The ball is placed at a spot 36 feet from the goal; the goalkeeper may not move until the ball is struck. The international soccer goal is 8 feet high and 24 feet wide. In an analysis of penalty kicks taken in the World Cup between 1982 and 1998, it was determined that 211 such kicks were taken, with 161 successfully made: Success rate = 161/211 = 76.3%.

Further analysis revealed that the goalkeeper during the penalty dove to the side to which the ball was directed (the correct side) 63% of the time. Also, 41% of all successful goals were scored within 6.2 feet of the goalkeeper's initial position. By studying the tendencies of an opponent (that is, whether the kicker tends to kick the ball in a particular direction, or to a particular part of the net on penalties), a soccer goalkeeper can increase the chance of saving a penalty.

GOLF TECHNOLOGY

In recent years, there has been considerable public debate about the technology of golf clubs, and the ability of golfers to hit a golf ball farther than ever. One key area of debate has centered on the construction of the driver, the club used to generate the greatest distance. World golf regulatory bodies have imposed rules with respect to the construction of the driver, based on the principles of coefficient of restitution (COR). As noted in the baseball math segment, COR is the ratio of the speed of

an object measured before and after a collision with a fixed object, such as a wall. It is impossible to have an object speed after such a collision that is greater than the object speed prior to collision. The higher the COR (that is, the closer the COR is to 1.0), the faster the object is expected to move after collision. If the object, in this case a golf ball, were struck by a very bouncy, trampoline-type surface on the face of the golf club, one would expect it to travel farther than if struck by a denser, less elastic material. Huge sums are spent each year by golf equipment manufacturers to design surfaces for clubs that create high COR values.

The current COR for a driver legal for use in international golf is 0.83, meaning that an object striking the material used on the driver's surface at 100 miles per hour would expect to rebound at a speed of 83 miles per hour.

FOOTBALL—HOW FAR WAS THE PASS THROWN?

There are innumerable circumstances where mathematical principles can be used to assist in assessing performance. For example, to determine how far a quarterback actually threw a pass, consider the following: The quarterback takes the snap from the center and drops back to pass. The pass is delivered and is caught by a receiver 20 yards from the quarterback's position. The ball was thrown at an angle of 30° from the line of the hash mark on the field. What was the actual down-field gain for this pass and catch? The distance thrown down field is represented by d: $d = 20 \times \sin 30° = 20(0.5)$; $d =$ 10 yards.

MONEY IN SPORT—CAPOLOGY 101

To even the casual observer of the modern sporting world, the media coverage of teams and competitions is seemingly fixated with the financial aspects of sport. Player contracts, television contracts, the sale of professional franchises for huge sums of money, ticket prices to attend events, all these issues have captivated the sporting public to an ever-increasing degree.

The salary cap is a financial tool in place in a number of professional sports. By definition, a salary cap is a prescribed limit placed upon how much money individual athletes may earn from their playing contract, and the salary cap is also a limit as to how much a team may collectively pay its roster of players. NFL football and NBA basketball are the two best North American examples of leagues with a salary cap in place.

As with other examples of mathematics in sport, the expression of the salary cap in a sport as a finite number may appear simple; the calculation and the impact of the salary cap on different aspects of team organization and player transactions is often very complicated. The salary cap and its rules as employed in various sports have created a species of sports administrator commonly referred to as the team "capologist," an expert with respect to the interpretation of salary cap rules made by the league in question. The capologist will assist the team management in determining whether, from a financial perspective, certain types of player transactions comply with the salary cap rules.

The salary cap is generally intended to create two important results for a professional team. One, the owners of the team will have a measure of cost certainty, in that they will know that in the given season, the team's player payroll will not (or should not) exceed the cap limit. For example, if an NBA basketball team is said to have a salary cap of $82 million, the team payroll, in theory, may not exceed this amount, and the team must budget accordingly. Two, the level of the salary cap will impact decisions that the team may wish to make concerning trades and other acquisitions of players. As noted, the salary cap may be set out in a finite number.

The NFL salary cap structure is a complicated calculation, taking into account numerous factors. For the sake of illustrating the function of the salary cap and its impact upon team personnel decisions, the example is the amount of the salary cap. This figure is calculated using 64% of the team's "defined gross revenue" calculated from the previous year. For the purpose of this calculation, such revenues will include the team's share of the league television revenues, stadium ticket sales, merchandise sales, and related revenues generated by the games themselves for each team.

Therefore, in an imaginary NFL season, if a team had defined gross revenues of $120 million, the salary cap = $120,000,000 × 0.064 = 76,800,000; the salary cap for the next season would be $76.8 million.

By salary cap rules, the top 51 players' contracts on a team are included for the purposes of the salary cap. The amount of the contract is defined by both its face value, for example, if a player has a contract worth $10 million over a four-year period, as well as any bonuses that the individual contract may provide. All bonuses are prorated for the purpose of a salary cap calculation over the four years of the sample contract.

In the sample, if the player had a $10 million contract over four years, and a $1 million bonus he received upon signing the contract, for the purposes of the salary cap, the contract is expressed as $10 million / four years = $2,500,000/year (salary component); $1 million / four

years = $250,000/year (bonus component); net salary = $2,750,000/year. The player salary is treated as $2.75 million against the salary cap total of $76.8 million. This calculation will be made with respect to all 51 current contracts.

Assume that the total player salaries are $71.2 million. The total available monies with which to sign other players to contracts is $5.6 million for the coming season, subject to releasing or otherwise terminating any existing contracts to create a greater cushion under the salary cap limit of $76.8 million.

How does the salary cap work if a team wishes to acquire a player beyond their means? In the example, the available money for player acquisitions is $5.6 million. The team finished the previous season at eight wins and eight losses, and it did not qualify for the postseason playoffs. The head coach and the general manager believe that a certain wide receiver, who is not under contract to any team and is therefore a free agent, would be a player who might take the team that extra step needed to make the league playoffs in the coming year.

This wide receiver is an elite player, and he is expected to command a salary of $10 million per season, and he will command a contract of four seasons. Can the sample team with only $5.6 million left in its salary cap sign this player? The capology options include:

- No bid for this player: The current roster, subject to other contingencies such as injury, remains intact. (In a salary cap, where a player has been injured, they remain in receipt of their salary for the life of the contract, all counted in some fashion against the salary cap.)
- Sign the elite player at $10 million per season for four seasons. To get "under" the salary cap in this example, the team would be required to cut other players whose salaries total $4.4 million for the coming season ($10 million in new salary, less the available $5.6 million). The team in this scenario would be required to assess whether the benefit to the team in terms of performance was worth the loss of other players; further, the variable of injury for the new player would be considered.
- Sign the elite player, but structure the $10 million salary in year one of the four years as follows: Agree that the contract will be a $20 million bonus, and $20 million in salary over the following three years. The bonus is prorated over four years, meaning only $5 million would count against the salary cap this coming season. As $5.6 million is available as room under the team's cap, the bonus/deferred salary structure works, at least for the first year. The team will have to assess how it deals with this contract in each successive year, as it will be required to count this player's salary contract in year two as Bonus = $5 million (25% calculated over four-year period) and Salary = $20 million obligation now payable over three years.

This math application in essence borrows from the team's future to pay for the present needs of the team. In the realm of the salary cap, the best interests on the team on the field and the best financial interest of the team do not always exist in harmony.

The more involved the mathematical equations dealing with salary cap, the less important are the players themselves. Further, it is a reasonable presumption that the greater the room available to a professional sports franchise in its salary cap, the greater potential profits to the ownership of the franchise.

Some salary caps have a punitive component for those teams that breach the salary cap rule; these penalties are often referred to as a luxury tax. The premise behind these measures is that the richer franchises that exceed the salary cap limits will pay monies back into the general funds of the league, which are then distributed among the franchisees that abided by the salary cap rules.

In the NBA, the tax on the individual player salary that broke the cap ceiling is 10%. The team is also obligated in general terms to pay a 10% team tax on its payroll that is in excess of the cap. There are a multitude of exemptions and qualifications; the bottom line for the owner is, are they prepared to exceed the salary cap and pay the penalties imposed if they get a team that might win a championship?

MATH AND SPORTS WAGERING

Team sports wagering has grown from its clandestine roots in taverns and clubs to a multi-billion dollar enterprise that includes private bookmakers and state-run sport bets. All forms of sport gambling have a mathematical basis, rooted in the concepts of probability and understanding the statistics relied upon by odds makers to establish betting systems. There are a number of different types of wagers available, each generally involving a different math principle:

- Straight bet: This is a wager placed on the final outcome of an event. For example, if a team is chosen as the winner and does win, the successful bettor gets a return on their money 1:1. If $100 were wagered on the team, the winner recoups his initial bet, plus $100.
- Odds: As with the straight bet, the wager is with respect to the final outcome, with the odds, or the

Key Terms

Average: A number that expresses a set of numbers as a single quantity that is the sum of the numbers divided by the number of numbers in the set.

Odds: A shorthand method for expressing probabilities of particular events. The probability of one particular event occurring out of six possible events would be 1 in 6, also expressed as 1:6 or in fractional form as 1/6.

Percentage: From the Latin term *per centum* meaning per hundred, a special type of ratio in which the second value is 100; used to represent the amount present with respect to the whole. Expressed as a percentage, the ratio times 100 (e.g., 78/100 = .78 and so .78 × 100 = 78%).

Statistics: Branch of mathematics devoted to the collection, compilation, display, and interpretation of numerical data. In general, the field can be divided into two major subgroups, descriptive statistics and inferential statistics. The former subject deals primarily with the accumulation and presentation of numerical data, while the latter focuses on predictions.

probability, of the event added to the wager. For example, as in the earlier example, if the team were not likely to beat the opponent, the odds of such an event occurring might be as remote as 10:1 against, meaning that it is stated to be 10 times more likely that the team will lose than win. If $100 were wagered on 10 to 1 odds, and the team were successful, the successful bettor would again recoup the initial $100 wagered, plus 10 × 100, or $1,000.

- Point spread (also referred to as the line and other terms): This variation in sports betting is very popular in sports such as football and basketball. The nature of the point spread in any given game is typically calculated by professional gambling organizations, and published in major media. The bettor does not wager necessarily on the best team, but the wager is with respect to the difference in points between the team's scores at the end of the game. For example, Team A and Team B are NFL football teams scheduled to play on a Sunday afternoon. The professional gambling organization reviews the teams' records, injury situation, home field advantage, and the play of each team to date, and determines that "Team A is a 5-point favorite," which means that the gambling organization believes that Team A will beat Team B by 5 points or more. The organization will then take bets on the outcome of this game using that 5 points, referred to as the spread, as its betting standard for that game. The results in this type of bet for a bettor placing $100 on Team A are that Team A must win by 5 points or more. If Team A wins by 5 points exactly, the result is referred to as a "push": the bettor gets his $100 back, less the fee charged by the gambling house, 10%. Another result is for a bettor who places $100 on Team B. Because Team A is favored by 5 points, this bet will succeed if either Team B wins altogether, or Team B loses by 5 points or less. As with the straight bets, these wagers pay on a 1:1 ratio, less the 10% customarily charged by the betting establishment.
- Over/under: This bet and its variations are based upon the total number of points scored in a game, including any overtime played, by both teams; the win or loss of the game itself is not relevant. For example, in a basketball game, the wagering line would be established as 176 points, wagers invited as being over and under the mark. If a wager is successful in predicting whether the teams were a total over or under the line, the return is again a 1:1 ratio to the money wagered.
- Parlay: This form of wagering permits the bettor to gamble on two or more games in one wager. The bettor must be correct in all of the individual wagers to claim the entire bet. The reward multiplies in parlay betting, as does the risk of missing out on one wager in the sequence:In three-game parlay, Game has 12.7:1 odds; Game 2 has 3.3:1 odds; Game 3 has 1.9:1 odds. On a $5.00 wager on this three-game parlay, the return if each team selected were successful would be 2.7 × 3.3 × 1.9 = 16.93; $5 × 16.93 = $86.45. As is illustrated, a return of almost 17 times the initial $5 wager would be a successful gambler's reward in this scenario; a loss of any of the three games would mean the bettor would lose the entire parlay.
- Future event: It is common for both North American and world sporting events to be the subject of odds

posted by various professional gambling agencies. For example, in the lead-up to the World Cup of Soccer, every team will be the subject of odds of winning the quadrennial championship; a perennial soccer power like Brazil might be listed at 3 to 1 odds, while a traditionally less successful nation, such as Saudi Arabia or Japan, will be listed at more dramatic numbers such as 350 to 1. Wagers are typically binding at the odds quoted, no matter what might happen to the subject team in the period between the date of the wager and the date of the event. For example, if Brazil's best scorer and best goaltender were injured, the actual odds quoted for Brazil might be quite higher at the start of the championships; the wager would remain payable at the initial 3 to 1 odds.

Where to Learn More

Books

Adair, Robert K. *The Physics of Baseball,* 3rd ed. New York: Perennial, 2002.

Holland, Bart K. *What are the Chances? Voodoo Deaths, Office Gossip and Other Adventures in Probability.* Baltimore, MD: Johns Hopkins University Press, 2002.

James, Bill. *Baseball Abstract,* Revised ed. New York: The Free Press, 2001.

Periodicals

Klarneich, Erica. "Toss Out the Toss Up: Bias in Heads or Tails," *Science News*, February 28, 2004.

Postrel, Virginia. "Strategies on Fourth Down, From a Mathematical Point of View," *New York Times*, September 9, 2002.

Square and Cube Roots

Overview

Finding the square and cube roots of a number are amongst the oldest and most basic mathematical operations. A number, when multiplied by itself, equals a number called its square. For example, nine is the square of three. The square root of a number is the number that when multiplied by itself, equals the original number. For example, three is the square root of nine. The cube root is the same concept, but the cube root must be multiplied three times to yield the original number. These two concepts get their names from the relationship they have with the area of a square and the volume of a cube.

In our three dimensional world, lines that have one dimension, squares that have two dimensions, and cubes that have three dimensions form the basic shapes that mankind uses to build models of the world. The square and cube of a number, and their inverses the square and cube roots, allow us to relate the length of a line to the area of a two-dimensional square or the volume of three-dimensional cube respectively.

Examples of the square and cube roots will be found in any area of design where a model of an object will need to be conceptualized before the object can be built, for example in the architect's plans for a new house or the maps for the construction of roads, or the blueprints of an aircraft. During the design phase, whenever areas and volumes need to be manipulated, the square and cube roots would be used to calculate these quantities.

Fundamental Mathematical Concepts and Terms

The definition of the square root is a number that when multiplied by itself, will yield the original number. As an example, again consider the value 9. It has a square root of 3, so $3 \times 3 = 9$. The value 9 is called the square of 3. The cube root is similar, but now the value that has to be multiplied is multiplied by itself three times, for example, the cube root of 8 is 2, so $2 \times 2 \times 2 = 8$ and the value 8 is called the cube of 2.

The names square and cube root come from their relation with these shapes. Consider a square, where each side has an equal length; if you know the area of the square, the square root will give you the length of one side. Since all the sides are an equal length, you have found the length of them all. The area may be some square land where you want to know how much fencing is needed to mark the edge of your land. If the area is 100 square meters then the length of one edge is 10 meters. As

there are four edges to the square, you will need to buy 40 meters of fencing.

The cubed root comes from the same idea. Imagine a wooden cube, where each edge is again exactly the same length. If we know the volume of this cube, the cube root will give us the length of one of the edges; since it is a cube, we know the length of all the edges. For example, an architect has calculated that his building will need a foundation with 1000 cubic meters of cement to hold the weight of the structure safely. The cube root of 1000 is 10, so the builders will know that by marking a 10 by 10 meter square out on the floor and digging down 10 meters this hole will be the right size for the cement.

NAMES AND CONVENTIONS

In mathematical text the radical symbol is used to indicate a root of a number. The square root is written as $\sqrt{9} = 3$.

To indicate roots or higher than the square root, for example the cubed root, the number of the root is entered into the top left part of this symbol. For example the cubed root is written as $\sqrt[3]{8} = 2$.

This notation was developed over a period of about 100 years. The right hand slash and line above the numbers first appeared in 1525 in the first German algebra book, *Die Coss*, by Christoff Rudolff (1499–1545). It is thought that the notation of adding the number 3 for a cube and numbers for higher roots as a symbol to the top left of the radical was first suggested by the Western philosopher, physicist, and mathematician René Descartes (1596–1650). The addition of the "vee" to the left side of the symbol is thought to have been developed in 1629 by Albert Girard (1595–1632), a French mathematician who had some of the first thoughts on the fundamental theorem of algebra.

The name root comes from a relationship with a family of equations called polynomials, these equations contain all the powers of a variable x in an infinite series and have the form, $y = a + bx + cx^2 + dx^3 + ex^4 \ldots$ and so on, forever. All the letters on the right hand side of the equals sign, apart from the x, can have any values we want. Setting a value to zero will eliminate that term in the series.

A Brief History of Discovery and Development

In ancient times numbers held a deep religious and spiritual significance. Mathematics was heavily based on geometry, philosophy, and religion. Early thinkers about the nature of geometry saw lines and other geometrical shapes as the fundamental and logical building blocks of the heavens and Earth. The idea that nature could always be expressed with lines and shapes lead to the development of Pythagoras' famous proof for triangles, a relation that uses the square root to calculate the final answer.

Pythagoras of Samos (c. 500 B.C.), was an extremely important figure in the history of mathematics. Pythagoras was an ancient Greek scholar who traveled extensively throughout his life. He founded a school of thought that had many followers. The society was extremely secretive but was based on philosophy and mathematics. The school admitted women as well as men to follow a strict lifestyle of thought and practice of mathematics.

Pythagoras' proof is for a triangle with one right angle and it relates the length of the longest side to the lengths of the other two sides. In the modern era, the proof is included in school textbooks and so it is hard for us to understand the deep impact on their way of life that this new method of logical thinking had on our ancestors. The proof—and knowledge of mathematics in general—were venerated as sacred secrets.

Today, Pythagoras' proof is learned as a formula with symbols, but this system of thinking would not have been known to its founder. Moreover, the proof that Pythagoras found was based purely on geometry. Legend has it that a philosopher of Pythagoras's society, called Hippasus, made the discovery at sea that if the two shorter sides of the triangle are set to 1 unit of length, then the result for the longer sided is an irrational number when the square root is taken. This special number could never be drawn with geometry and the legend goes that the other Pythagoreans were so shocked at this discovery that they threw him overboard to drown him and so keep his discovery a secret.

There is another important property of taking roots of numbers that was not understood until English physicist and mathematician Sir Isaac Newton's (1642–1727) time: the concept of taking the root of a negative number. If you try this on a calculator it will most likely give you an error. However, it was shown that it is possible to extend our number system to deal with taking the root of a negative number if we add a new number, given the symbol, i, in mathematics. This opened a whole new world of algebra that mathematicians call complex numbers and allows solutions to be found for problems that had previously been thought impossible.

From a practical viewpoint, this development affected almost every area of modern physics, which relies on complex numbers in some form or another. Some examples of their usage are found in electromagnetism, which gave us television, radio, and quantum mechanics,

which gave us, among many other things, computers and modern medical imaging techniques.

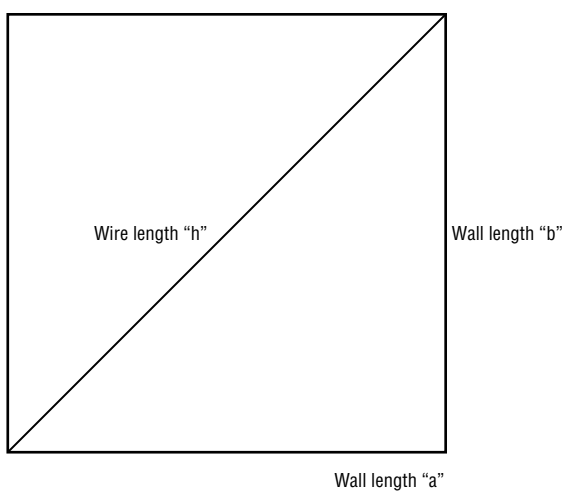
PYTHAGOREAN THEOREM

Using just pure geometry, Pythagoras is famous for proving that, for a right angled triangle, the square of the lengths of the longest side, called the hypotenuse, is equal to the sum of the squares of the other two sides. This rather long sentence is much easier to follow if it is written as an equation: $h^2 = a^2 + b^2$.

In this equation, the letter h is the length of the hypotenuse and a and b are the lengths of the other two sides. As this equation has only squared terms, we must take the square root if we want to find the actual length of h.

For example, in a rectangular room, how long would a wire have to be if it was to be run in a straight line, across the floor, from the back, left hand corner, to the front, right hand corner? The room is full of furniture and it would be impossible to just measure the distance with a tape measure. However, we notice that the walls and the wire form a triangle pattern. Each wall is at right angles and lengths of the walls form the shorter two sides of the right angle triangle. The wire, running across the room, forms that longer side, the hypotenuse.

Finding the length of a wire



$$h^2 = a^2 + b^2$$
$$h = \sqrt{h^2} = \sqrt{a^2 + b^2}$$

One wall is 3 meters long, and the other is 7 meters, so: $h^2 = 3 \times 3 + 4 \times 4 = 25$. So the length of the wire is given by the square root of 25 as 5 meters long.

HIPPASUS' FATAL DISCOVERY

How long is the wire in the previous example if we have a room where each wall is just 1 meter long? $h^2 = 1 \times 1 + 1 \times 1 = 2$. Now take the square root of 2 to find 1.4142136.

In fact the digits of this number go on forever. It is a member of the family of numbers called irrational numbers. These numbers have the property that the fractional part of the digits continue forever and never repeat the same pattern. From the practical perspective of installing our wire, this is no problem as we would simply round up the length. However, in the exact world of mathematics the consequences are much more dramatic. Due to the fractional part having an infinite nature, it cannot be expressed as a ratio of integer values (a fraction).

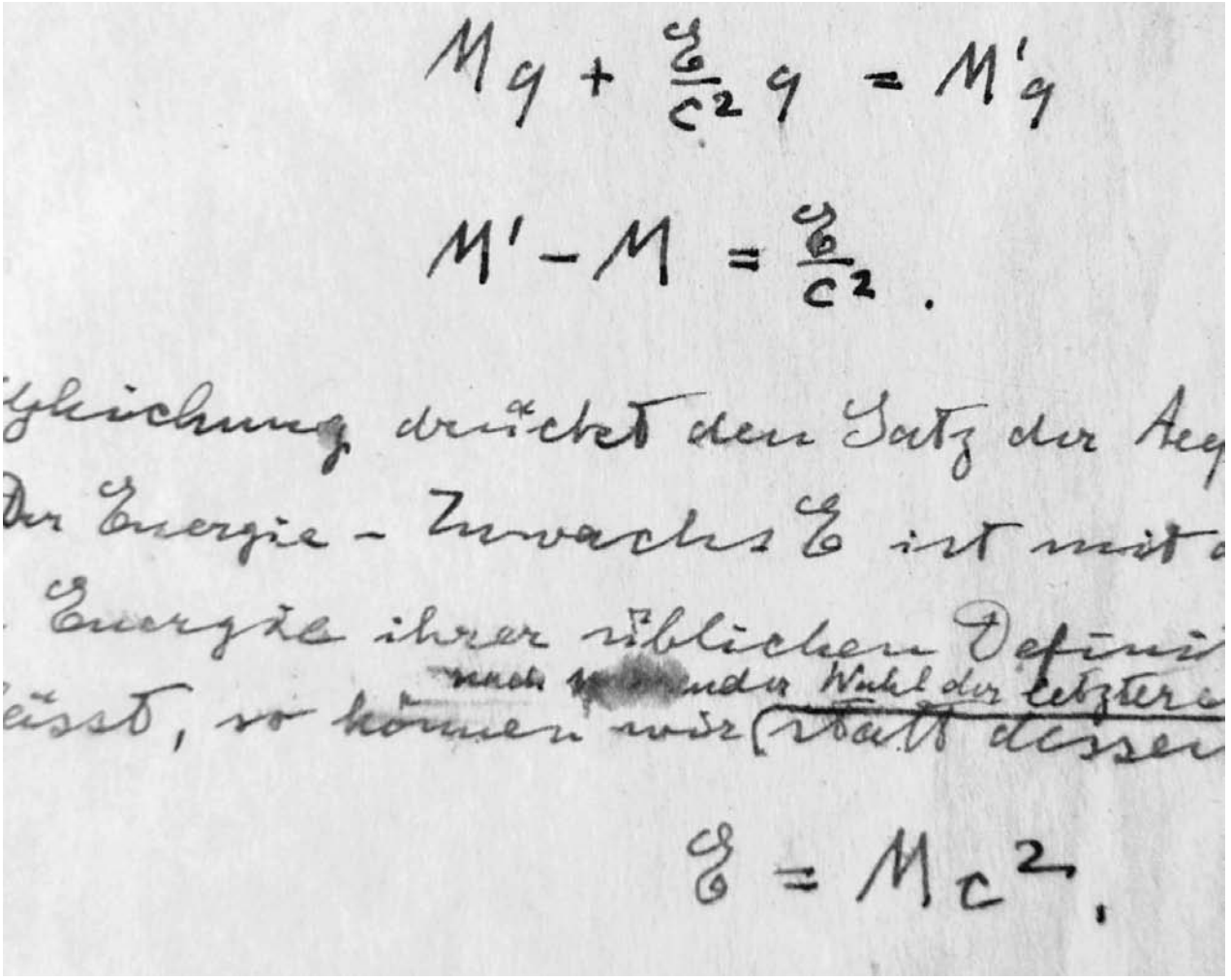
What is even stranger is that we have made this length in something that is a perfectly reasonable and real geometric shape, a square box with sides equal to 1 meter. In this case, what exactly does the length of the line from one corner to the opposite corner of the box "mean"? Something that at first glance would seem child's play to measure is soon found to be impossible. No matter what we do, the length, given by the square root of 2, will always be wrong to some degree if we try to give it an exact value. In the legend of the death of Hippasus at the hands of his fellow Pythagoreans, it was the discovery of this anomaly that shattered the idea that the Heavens and Earth could be expressed totally and completely by lengths and their ratios.

Real-life Applications

ARCHITECTURE

The knowledge that some lengths are related with squared ratios has been known since Egyptian times, even though they would not have known the proof. Examples of this include the lengths 3, 4, 5, which are related by Pythagoras' theorem and are thought to be found in the construction of the Egyptian pyramids.

Today, squared and cubed roots are used in construction and design. If you were to design a car you might wish to change the volume of the driver's compartment. A modern three-dimensional (3D) design would be stored, as a wire frame model, in the memory of a computer. A computer program will divide the 3D space into thousands of tiny cubes, a job that is easy for a computer to do. Next, a program is run that counts the number of cubes within the driver's compartment and returns a value. The total volume is equal to the number of cubes found in the compartment, multiplied by the



$$Mq + \frac{\mathcal{E}_0}{c^2} q = M' q$$

$$M' - M = \frac{\mathcal{E}_0}{c^2}.$$

Gleichung drückt den Satz der Äq
der Energie-Zuwachs $\mathcal{E}_0$ ist mit
Energie ihrer üblichen Defini
nach [illegible] der Wahl der letzteren
lässt, so können wir (statt dessen

$$\mathcal{E}_0 = Mc^2.$$

This paper, written in 1946, was written by Albert Einstein. He explains how he derived the formula $E = Mc^2$, a consequence of his Special Theory of Relativity, first published in 1905. The formula specifies that c (the speed of light) is squared. AP/WIDE WORLD PHOTOS. REPRODUCED BY PERMISSION.

volume of one cube. The one cube is called the unit cube and has real dimension; this allows us to make modifications to the actual size of the 3D wire frame without altering the wire frame itself.

To change the volume of the compartment, you change the volume of the unit cube. The amount that you would need to scale the sides of the unit cube is found by taking the cubed root of the original volume

NAVIGATION

The use of Pythagoras' theorem allows distance to be calculated on maps using coordinate systems. A coordinate system is a grid-like structure that is used as reference for points on the map's surface. Lines between one point and another form vectors and the calculation of lengths of vectors requires the use of square roots. Vectors can also be used to map velocity, a combination of speed and direction. These systems are used on land by the military, at sea by the navy and shipping firms, and in the air by aircraft, to plan and negotiate the terrain they are moving over. As an example, if two ships are moving perpendicular to each other, i.e, at 90 degrees to each other, and one ship is traveling at 3 knots and the other at 4 knots, using Pythagoras' relation, the navigators on the deck of each ship would measure the speed of the other as moving away from them at 5 knots.

SPORT

Football pitches, tennis courts, race tracks, and swimming pools are some examples of areas used by professional

Key Terms

Cubed root: The relation of the volume of a cube to one of its edges.

Root: The solutions of a polynomial equation, of which the square and cube root are special cases.

sports people that need to be accurately measured if the events are to be considered fair. The areas to be surveyed and locations of the various markings must be set down. The process of surveying these areas requires the use of roots in the calculations of various lengths for the markings

STOCK MARKETS

Many of the transactions used in stock markets use statistics to estimate the market trends and the best times to buy and sell stocks and shares. These calculations will often use something called the standard deviation, a measure of the spread of random events, and will give the traders some idea of the accuracy of their estimates. This calculation will require the use of roots.

Another occurrence of the root comes when the errors of predictive models are calculated. Models used to predict the stock market or anything else will have some sort of error depending on the accuracy of the data fed into it. If the error is much smaller than the size of the result, then the result can be trusted.

For example, if your model suggests that you buy gold next Wednesday, within an error of one hour, this is fine, but if the error is ten years then the it would be foolish to trust the result. As there may be many sources of error they will all have to be accounted for they need to be combined to give a final overall error. This technique is well defined in statistics, which requires the use of the square root.

Potential Applications

GLOBAL ECONOMICS

As global finance becomes more sophisticated, mathematicians and economists investigate the patterns of these transactions and look for relationships that will indicate the growth and decline of large groups of companies or even countries. It has only been recently, with the large scale computing and the application of a number of areas of science to economics that such models have come into use.

Successful interpretation of these trends, and new ideas and concepts in understanding the trends, are vital to the future development and stability of corporations and governments. This science, macroeconomics, is statistical in nature and allows predictions of important economic indicators such as inflation, interest rates, and the prices of materials. The use of squared and cubed roots in making these judgments incorporates fundamental formulas of probability and statistics that rely on square and cube roots.

Where to Learn More

Web sites

Wolfram. MathWorld. <http://mathworld.wolfram.com/> (February 1, 2005).

Statistics

Overview

Statistics is the branch of applied mathematics concerned with characterization of populations by the collection and analysis of data. Its applications are broad and diverse. Politicians rely on statistical polls to learn how their constituents feel about issues; medical researchers analyze the statistics of clinical trials to decide if new medicines will be safe for the general public; and insurance companies collect statistics about automobile accidents and natural disasters to help them set rates. Baseball fans immerse themselves in statistics that range from slugging percentages to earned run averages. Nervous travelers comfort themselves by reminding themselves that, statistically speaking, it is safer to travel in a commercial airliner than in an automobile. Students preparing for college fret over grade point averages and standardized test score percentiles. In short, almost every facet of daily life involves statistics to one degree or another.

Fundamental Mathematical Concepts and Terms

POPULATIONS AND SAMPLES

A statistic is a numerical measure that characterizes some aspect of a population or group of values known as random variables. They are random variables because the outcome of any single measurement, trial, or experiment involving them cannot be known ahead of time. The weight of men and women, for example, is a random variable because it is impossible to pick a person at random and know his or her weight before he or she steps on a scale. Random variables are discrete if they can take on only a finite number of values (for example, the result of a coin toss or the number of floods occurring in a century) and continuous if they can take on an infinite number of values (for example, length or height).

In some cases the populations are finite, for example the students in a classroom or the citizens of a country. While it may be impractical to do so if the population is large, a statistician can in theory measure each member of a finite population. For example, it is possible to measure the height of every student attending a particular school because the population is finite. In other cases, especially those related to the outcome of scientific experiments or measurements, the populations are infinite and it is impossible to measure every possible value. An oceanographer who wants to determine the salt content of sea water using an electronic probe is faced with an infinite population because there are an infinite number of places where he or she could place the probe.

In many practical situations, the underlying objective of statistics is to make inferences about the characteristics of a large finite or infinite population by carefully selecting and measuring a small sample or subset of the population. A political pollster, for example, may infer the likely outcome of a national election by asking a sample of a few hundred carefully chosen voters which candidate they prefer. An environmental scientist may collect only a few dozen samples in order to determine whether the soil or water beneath an abandoned factory is contaminated. In both cases it would have been impractical or impossible to analyze each member of the population, especially because the number of possible samples that could be collected is infinite. So, representative samples are chosen and statistics are calculated to draw conclusions about the population. Statistics that are calculated from measurements of an entire finite population are known as population statistics, whereas those that are based on a sample of either a finite or infinite population are known as sample statistics.

Because sample statistics are used to make inferences about populations, it is essential that the samples are representative of the population. If the objective of a study is to calculate average income, then it would be misrepresentative to poll only shoppers at a yacht brokerage because people who can afford yachts probably have incomes that are higher than average. By the same token, it would be just as misrepresentative to ask people waiting in line to file unemployment claims, because their incomes may generally be lower than average. Therefore, real world applications of statistics demand that considerable attention be given to experimental designs and sampling strategies if the statistical results are to be reliable.

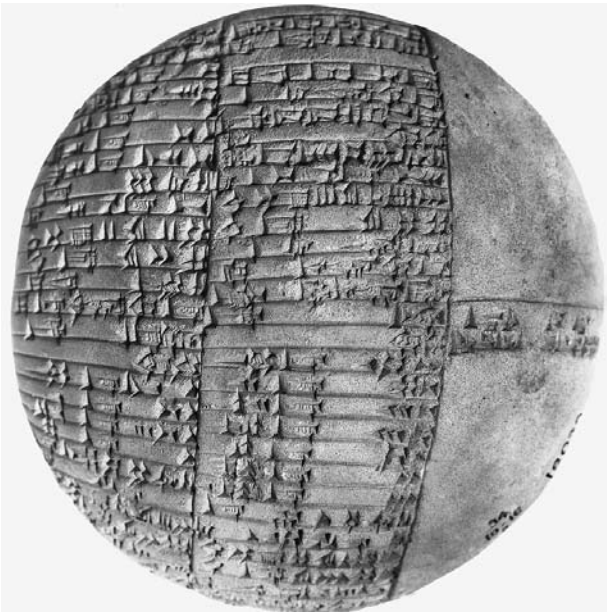
One way to obtain a representative sample is to select members of the population at random. In simple random sampling, each member of the population has an equal chance of being selected or measured and there is no predefined sampling pattern. Random sampling is often accomplished using a computer program that generates random numbers or by referring to published random number tables. It is impossible to generate truly random numbers using a computer program, because the program itself must have some underlying structure or pattern. Mathematicians have been able to develop methods or algorithms, however, which generate nearly random numbers that suffice for most practical applications. To select a random sample of 100 people attending a sporting event, a statistician might assign a number to each seat in the stadium or arena. Then, he or she would generate 100 random integers and the people in the seats corresponding to those 100 numbers would comprise the random sample. Likewise, a scientist interested in measuring the soil nutrients in a farmer's field might divide the field using a grid of north-south and east-west imaginary lines. If the objective were to sample the soil at 20 random locations, the scientist would then use 40 random numbers to generate 20 pairs of north-south and east-west coordinates. One sample would be taken at each of the 20 locations specified by the coordinates.

Although simple random sampling works well for homogeneous populations, it may not produce truly random samples of heterogeneous populations that consist of distinct sub-populations or categories. In such cases, stratified random sampling provides more representative samples. The first step in stratified random sampling is to define the sub-populations. In a political poll, the sub-populations might be registered Democrats, Republicans, and Independents. In a marketing survey, the sub-populations might be defined in terms of age, sex, and income. Each sub-population is randomly sampled and the results are weighted so that they are proportionate to the relative size of each sub-population. Thus, stratified random sampling provides results that characterize each sub-population and the population in general, which the contribution of each sub-population proportional to its size.

PROBABILITY

It is possible to use basic statistical results without reference to the concept of probability. A diehard baseball fan, for example, can compare Babe Ruth's lifetime batting average of 0.342 to Hank Aaron's lifetime batting average of 0.305 and argue passionately that Ruth was the better hitter of the two. Batting averages are statistics, one is clearly larger than the other, and there is no need to worry about the nature of probability.

Unlike simple comparisons of batting averages, real life applications of statistics are in most cases closely tied to the concept of probability. The type of probability that is most often taught in basic statistics courses is known as relative frequency probability (or just frequency probability), and those who advocate this definition are known as frequentists. Relative frequency probability is defined as the number of times an event has occurred divided by the number of trials conducted or observations made, where the number of trials or observations is large. Flip a coin many times and the results should be very close to 500 heads and 500 tails, so the relative frequency is $500 \div 1{,}000 = 0.5$, or 50%. All other things being equal, therefore, the probability of obtaining a head with the next toss is 50%. A slightly more complicated example might involve the measurement of a quantity that has an infinite number of possible outcomes, for example weight. If each



This tablet displays ancient Sumerian measurements and statistics (ca. 2400 B.C.). BETTMANN/CORBIS.

of 1,000 students in a high school were weighed, and 100 of them weighed between 140 and 150 pounds, then the relative frequency of a weight in that interval would be 100/1,000 = 0.1, or 10%. Therefore, the probability that a student selected at random would weigh between 140 and 150 pounds is 0.1. The determination of values of a random variable, in this case the weights of students in a school, by repeated measurement produces an empirical probability distribution.

Mathematicians have devised a number of theoretical probability distributions that play an important role in statistics, the best known of which is the normal, or Gaussian, distribution. Named after the mathematician Karl Friedrich Gauss (1777–1855), the normal distribution is defined by a probability density function that follows a distinctive bell shaped curve. Continuous random variables following a normal distribution are more likely to have values near the peak of the curve than near the ends. In many situations, it is the logarithms of values, not the values themselves that follow a normal distribution. In this case the distribution is said to be lognormal. Another example of a widely used theoretical probability distribution is the uniform distribution, which is defined by minimum and maximum values. Each value in a uniform distribution has an equal probability of occurrence. The binomial distribution applies to discrete random variables.

Although the normal (and lognormal), uniform, and binomial distributions are among the most common probability distributions, there are many specialized distributions that are particularly well-suited for specific problems. The Pareto distribution, for example, is named after the Italian economist Vilfredo Pareto (1848–1923) and is used in many statistical problems that consist of many small values and relatively few large values. It has found applications in studies of the distribution of wealth, the distribution of wind speeds, and the distribution of broken rock sizes encountered in construction and mining.

The great value of theoretical probability distributions, especially the normal distribution, is that they facilitate the use of rigorous mathematical tests that scientists can use to evaluate hypotheses and understand uncertainties in experimental data. For example, how likely is it that two samples were drawn from the same population? How certain are regulators that water quality meets government standards? How precisely must a product be manufactured to ensure that there is less than 1 defect in 1,000,000? How reliable are the results of a public opinion survey? The answers to these kinds of questions are more precise if the sample distribution follows a theoretical distribution and parametric statistical tests can be used. Therefore, one of the first steps in the statistical analysis of data is to determine whether the data are normally (or lognormally) distributed.

Statistics or statistical tests that are tied to a theoretical probability distribution are known as parametric. Those that are independent of any theoretical distribution are known as non-parametric.

MINIMUM, MAXIMUM, AND RANGE

The most fundamental statistics that can be calculated from a set of observations are its minimum value, maximum value, and range, which is the difference between maximum and minimum values. If the set of observations comprises the entire population, then the minimum and maximum will represent the true values. If the observations are only a sample of a larger population, however, the true or population minimum and maximum will be smaller and larger, respectively, than the sample minimum and maximum.

Consider the following list of values as an example: 8.95, 6.93, 11.07, 10.21, and 10.31. In order to calculate the range, first identify the minimum and maximum values in the list. In this case, as in most real life applications, the minimum and maximum values are not the first and last values. The minimum and maximum values in this example are 6.93 and 11.07, so the range is 11.07 − 6.93 = 4.14.

AVERAGE VALUES

An average is defined as a number that typifies or characterizes the general magnitude or size of a set of numbers. In statistics, there are several different types of averages known as the mean, median, and mode. The word average itself, however, does not have a formal statistical definition and is generally not used in statistical work.

The most common kind of average is the arithmetic mean, which is found by adding together all of the numbers in a lists and then dividing by the length of the list. Using the same list of numbers as in the previous section, the arithmetic mean is (8.95 + 6.93 + 11.07 + 10.21 + 10.31)/5 = 9.49. Another kind of mean, the geometric mean, is calculated using the logarithms of the values. The geometric mean is calculated as follows: First, find the logarithm of each number in the sample or population. For the example list of five values used above, the natural (base e = 2.7183) logarithms are: 2.19, 1.94, 2.40, 2.32, and 2.33. Second, calculate the mean of the logarithms, which is (2.19 + 1.94 + 2.40 + 2.32 + 2.33)/5 = 2.24. Finally, raise e to that power, or $e^{2.24} = 9.37$. Any base can be used to calculate the logarithms as long it is used consistently throughout the calculation. Statisticians sometimes refer to the arithmetic mean of a population as its expected value.

Another kind of average, the median, is the number that divides the sample or population into two subsets of equal size. If the list of numbers for which a median is to be calculated is of odd length, then the median is found by ordering or sorting the values from smallest to largest and selecting the middle value. If the list is of even length, the median is the arithmetic average of the two middle values of the sorted list. The sorted version of the example list from the previous paragraph is 6.93, 8.95, 10.21, 10.31, and 11.07. The length of the list is odd and the middle value is in position (5 + 1)/2 = 3, so the median is 10.21.

Although sorting is a trivial computation for a short list of numbers, sorting large lists can be time consuming and the development of fast sorting algorithms has been an important contribution to applied mathematics and computer science. To illustrate how a simple sorting algorithm works, compare the first two values of the sample data set from the previous paragraph, 8.95 and 6.93. The second value, 6.93, is smaller than the first value, 8.95, so the positions of the two values are switched. Next, the third value, 11.93, is compared to the first two. Because 11.93 is greater than both of the first two values, none of their positions in the list are switched. The fourth value, 10.21, is then compared. It is greater than the first two values, 9.93 and 8.95, but smaller than the third value, 11.93. Therefore, the positions of 10.21 and 11.93 are switched. The same procedure is repeated until each value in the list is compared and, if necessary, put into the correct position.

If a population follows a normal distribution or uniform distribution, its mean will be equal to its median. Another way of saying this is that the ratio of arithmetic mean to median is 1. If a population follows a lognormal distribution, however, the mean will be larger than the median. Scientists analyzing data often calculate the ratio of arithmetic mean to median as a simple preliminary method of determining whether the data are likely to follow a lognormal distribution. This is not a rigorous statistical method, though, and the preliminary result is often followed by more sophisticated calculations.

Astute readers will have noticed that the mean and median values calculated as examples in this section are not equal, but almost certainly will not know that the five numbers used in the calculations were selected at random from a normal distribution with an arithmetic mean of 10. If the five numbers represent a normal distribution, why are the mean and median different and why does neither of them equal 10? The answer is a consequence of the law of large numbers, which states that the difference between expected and calculated values decreases towards zero as the number of trials (in this case the number of randomly selected numbers) grows large. In other words, small sample sizes are likely to yield sample statistics that differ from the true population statistics. If the example calculations had been carried out using a list of 1,000 or 10,000 numbers, the sample arithmetic mean would have both been very close to 10. The corollary of this is that the reliability of sample statistics is generally proportional to the sample size. The larger the sample, the more likely it is that the sample statistics are accurate reflections of the underlying population statistics. In most practical applications, however, sample sizes are limited by the amount of money available to pay for the study (especially in cases where expensive laboratory tests must be conducted). The job of the practical statistician in many cases is to strike a balance between the desired accuracy of statistical results and the amount of money available to pay for them.

The third kind of average, the mode, is the most frequently occurring value in a sample or population. If no value occurs more than once, then the sample or population has no mode. If one value occurs more than any other, the data are said to be unimodal. Data can also be multimodal if more than one mode exists. For example, the list of values 3, 3, 4, 5, 6, 7, 7 has modes of 3 and 7.

MEASURES OF DISPERSION

Statistical measures of dispersion quantify the degree to which the values in a sample or population are clustered or dispersed around the mean. To illustrate the need for measures of dispersion, consider two samples. The first is 2, 3, 4, 5, 5, 6, 7, and 8. The second is 2, 3, 5, 5, 5, 5, 7, and 8. Both samples have identical minima, maxima, ranges, means, and medians, but the numbers comprising the second are more tightly grouped around the mean value of 5 than those in the first sample.

The most common measure of dispersion is the variance, which is based on the sum of squares of differences between the sample values and their mean. For the first set of example values in the previous paragraph, the mean is 5 and the sum of squared differences is $(2 - 5)^2 + (3 - 5)^2 + (4 - 5)^2 + (5 - 5)^2 + (5 - 5)^2 + (6 - 5)^2 + (7 - 5)^2 + (8 - 5)^2 = 28$. If the list of numbers represents an entire population, then the sum of squared differences is divided by the length of the list (in this case $n = 8$) to find the population variance of $28 / 8 = 3.5$. If the list of numbers represents a sample of a population, however, the sum is divided by one less than the number of values ($n - 1 = 7$) to find the sample variance of $28 / (8 - 1) = 4.0$. Repeating the calculation for the second sample, the result is $(2 - 5)^2 + (3 - 5)^2 + (5 - 5)^2 + (5 - 5)^2 + (5 - 5)^2 + (5 - 5)^2 + (7 - 5)^2 + (8 - 5)^2 = 26$. Depending on whether the result is for a population or sample, the variance is either $26/8 = 3.25$ or $26/(8 - 1) = 3.71$. Therefore, the variance of the second sample is smaller than that of the first even though the two samples have the same mean, minimum, and maximum values.

Because the variance is calculated from squared terms, the units of the values being calculated must also be squared. If the units of measurement are length (meters, for example), then the variance would be expressed in terms of length squared. The use of squared terms also means that variances will always be positive values.

The denominator used to calculate the sample variance is slightly larger than that used to calculate the population variance in order to account for the uncertainty or bias inherent any time that a sample is used to make inferences about a population. If the data set for which a variance is being calculated is the entire population, then the mean value used in the calculation is the population mean and the calculated variance is therefore unbiased. If the data set is a sample or subset of the population, though, the mean value is only an estimate of the population mean. Therefore, any subsequent calculations must take into account the fact that the use of the sample mean adds some bias to the results. This is accomplished by using a slightly smaller number ($n - 1$ rather than n) in the denominator to produce an unbiased estimate of the variance. The effect of dividing by $n - 1$ rather than n will decrease as the sample size becomes large, which reflects the fact that a variance calculated from a very large sample is a more accurate representation of the population variance than one calculated from a small sample.

Another commonly used measure of dispersion is the standard deviation, which is simply the square root of the variance. As such, standard deviations have units of plus or minus ($\pm$) the original units of measure. A variance of 4.0 meters^2 is therefore equivalent to a standard deviation of ± 2 meters. If the data being analyzed follow a normal distribution, then 68% of the values will fall within plus or minus one standard deviation of the mean, 95% will fall within two standard deviations of the mean, and 99.7% will fall within three standard deviations of the mean. If the data for which statistics are being calculated are measurements of error, for example the difference between the designed length and the actual length of an automobile part, then the standard deviation is often referred to as the root mean square or RMS error.

There are some situations in which the variance, and therefore the standard deviation, of a population is infinite. In such cases, attempts to calculate a variance will not converge on a single value as the sample size increases, and variances calculated using different samples of the same population will produce different results. It may still be possible, however, to calculate a statistic that is known as the average deviation, mean deviation, or mean absolute deviation. It is calculated in a manner similar to the variance, but the absolute values of each difference are used instead of their squares. The sum of absolute deviations of the sample 2, 3, 4, 5, 5, 6, 7, and 8 is thus $\text{Abs}(2 - 5) + \text{Abs}(3 - 5) + \text{Abs}(4 - 5) + \text{Abs}(5 - 5) + \text{Abs}(5 - 5) + \text{Abs}(5 - 5) + \text{Abs}(7 - 5) + \text{Abs}(28 - 5) = 12$, where Abs means "the absolute value of," and the average deviation is thus $12/8 = 1.5$.

Statisticians have largely avoided the average deviation for two reasons. First, it is difficult to work with absolute values when performing mathematical derivations. Second, the trick of dividing through by $n - 1$ rather than n to produce an unbiased estimate does not work nearly as well as with the variance. Therefore, statistics books do not contain alternative population and sample formulations for the average deviation. For the large data sets commonly encountered by many scientists and engineers, however, the difference between dividing by n and $n - 1$ is small enough to be inconsequential. Therefore, the average deviation is a statistic that has theoretical limitations but can be a useful practical tool for large data sets, and particularly those for which the variance is infinite.

CUMULATIVE FREQUENCIES AND QUANTILES

Cumulative frequency is closely related to relative frequency probability and has many applications in real life statistics. It is defined as the number of occurrences in a sample that are less than or equal to a specified value. If the cumulative frequency is divided by the number of data in a sample, it is, following from the relative frequency definition of probability, known as the relative cumulative frequency, cumulative probability, or plotting position. For a sample consisting of n data sorted from smallest to largest, the relative cumulative frequency of data point m is often calculated as $m/(n + 1)$. Consider this sample of five values: 19, 7, 20, 10, and 17. To calculate the relative cumulative frequency, first sort the list from smallest to largest to obtain 7, 10, 17, 19, 20. The relative cumulative frequency of 7, the first value in the list, is thus $1/(5 + 1) = 0.17$, or 17%. The relative cumulative frequency of 10, the second value in the list, is $2/(5 + 1) = 0.33$, or 33%. This procedure is repeated for each element in the list until a relative cumulative frequency of $5/(5 + 1) = 0.83$, or 83%, is obtained for the largest value. Thus, 17% of the values in the sample are less than or equal to 7 and 83% are less than or equal to 20. If the sample is representative of the population from which it was drawn, the same relative cumulative frequencies apply to the population. This approach also assumes that relative cumulative frequency is being calculated for a sample, not a population, because the formulation allows for the proportion 1/n of the values to fall below the smallest value in the list and 1/n of the values to fall above the largest value in the list. It is attributed to the Swedish engineer Waloddi Weibull (1887–1979), whose statistical formulations are often applied to analyze the sizes of events in sequences (for example, the sizes of yearly floods along a river).

Quantiles, sometimes known as n-tiles, are the values that correspond to particular relative cumulative frequency values. Using the data from the previous paragraph, the 0.17th is 7 and the 0.83rd quantile is 20. If the sample size is small, some quantiles will be undefined. For example, there is no 0.10^{th} in the list of five values used in the previous paragraph because none of the values has a relative cumulative frequency of 0.10. If it can be shown that the sample was drawn from a known theoretical distribution, such as a normal distribution, then statisticians can calculate the value that theoretically corresponds to a given quantile. The 0.25, 0.50, and 0.75 quantiles are often referred to as the first, second, and third quartiles, whereas the 0.01, 0.02, 0.03, 0.99 quantiles are often referred to as percentiles.

The Weibull formula, $m/(n + 1)$, is only one of several different ways to calculate the cumulative probability. In fact, the Weibull formula is somewhat arbitrary. The 1 was added to the denominator because data were at one time plotted on special graph paper, known as probability paper, which did not allow values of 0 or 1. This is because, strictly speaking, it is impossible for the probability of an event occurring to take on either of those values. Probabilities can come very close to 0 or 1, but never reach them. Another approach, known as Hazen's method, uses the formula $(m - ½)/n$ and is widely used in hydrologic studies. If it can be inferred that a sample follows a normal distribution, the quantiles can be calculated using a formula specifically designed for normal distributions. For most practical statistical problems there is usually very little difference between the values calculated using different methods.

CORRELATION AND CURVE FITTING

Correlation describes the degree to which two or more sets of measurements are related. For example, there is a general correlation between the height and weight of people (especially if they are of the same age, sex, and location). Correlation does not require a perfect relationship, but rather a degree of relationship or correspondence. It is possible that any given tall person weighs less than any given short person, but on average tall people will weigh more than short people.

Statisticians calculate correlation coefficients to express the degree to which two variables are correlated. The most common form of correlation coefficient is called the Pearson correlation coefficient, and is calculated using sums of mean deviations for each variable. It is almost always represented by r or R. Correlation coefficients can range from -1 to $+1$. A correlation coefficient of $r = 0$ represents a complete lack of correlation between two variables, and points plotted on a graph to represent the two variables will appear to be randomly located. Variables with correlation coefficients of $r = -1$ or $r = +1$ plot along a perfectly straight line, with the sign of the correlation coefficient indicating whether the slope of the line is negative or positive. In real life, most correlations fall somewhere in between these two extremes.

If two variables are correlated, it is often useful to express the correlation in terms of the equation for a straight line or curve representing the relationship. The simplest relationship is one in which the two variables are related by a straight line of the form $y = b + mx$. Because it is rare for variables to be perfectly correlated, the challenge is to find the equation for the line that fits data the best. There are several ways to do this, and all of them incorporate some way of minimizing the differences between the line and the data points. Regression is a

parametric, or distribution-dependent, procedure because it assumes that the differences to be minimized follow normal distributions. The general practice of finding the equation of the line that best represents the relationship between two correlated variables is known as regression or, more informally, curve fitting.

STATISTICAL HYPOTHESIS TESTING

In a previous example it was shown that the arithmetic mean of the numbers 8.95, 6.93, 11.07, 10.21, and 10.31 is 9.49. Could the numbers have been drawn at random from a normal distribution with a mean of 9 or less, even though the calculated sample mean is greater than 9? Possibilities such as this can be evaluated using statistical hypothesis tests, which are formulated in terms of a null hypothesis (commonly denoted as H_0) that can be rejected with a specified level of certainty. Statistical hypothesis tests can never prove that a hypothesis is true. They can only allow statisticians to reject null hypotheses with a specified level of confidence.

One common hypothesis test, the t-test, is used to compare mean values. It assumes that the values being used were selected at random from a normal distribution and that the variances associated with the means being compared are equal. It also takes into account the number of samples used to calculate the mean, because sample means calculated from a large number of values are more reliable than those calculated from a small number of values. The sample size is taken into account by using a probability distribution known as the t-distribution, which changes shape according to the number of samples. If the sample number is large, generally above 25 or 30, the t-distribution is virtually identical to the normal distribution.

To determine if the numbers 8.95, 6.93, 11.07, 10.21, and 10.31 are likely to have been drawn from a population with an arithmetic mean of 9 or less, first define a null hypothesis. In this case, the null hypothesis is that the arithmetic mean of the population from which the sample is drawn is less than or equal to 9. The result of the t-test, which can be performed by many computer programs, is a probability (p-value) of 0.27. This means that a person would be incorrect 27 out of 100 times if the population were repeatedly sampled and the null hypothesis rejected each time. Scientists often use a threshold (also known as a level of significance) of 0.05, so in this case the null hypothesis cannot be rejected because it is greater than either of those commonly used values. It can be tempting to interpret the failure to reject a null hypothesis at an 0.05 level of significance as a 0.95, or 95%, probability that the null hypothesis is true. But, this interpretation is inconsistent with the relative frequency definition of probability and should be avoided.

Similar tests can be conducted to compare the means of two samples (using a slightly different kind of t-test) or to compare the variances of two distributions (using an F-ratio test). In all cases, the tests are carefully structured so that the result is given as the probability of being incorrect if the null hypothesis is rejected.

CONFIDENCE INTERVALS

Another way to characterize the uncertainty associated with sample statistics is to calculate confidence intervals for the sample mean and variance. For the example of 8.95, 6.93, 11.07, 10.21, and 10.31, the confidence interval for the arithmetic mean at the 0.05 level of significance is 7.48 to 11.51. Calculation of the mean confidence interval relies on the t-distribution, so increased sample sizes will result in smaller confidence intervals. In other words, the larger the sample the more precisely the population mean can be estimated.

As above, the relative frequency definition of probability requires that this result be interpreted to mean that that true mean would be contained with the confidence interval 95 out of 100 times if samples of five were repeatedly drawn from the population. This is, strictly speaking, different than stating that there is a 95% probability that the population mean is between 7.48 and 11.51. The normal distribution from which the example values were drawn had a population mean of 10, so in this case the population mean did fall within the confidence interval. An analogous test can be performed to calculate confidence intervals for the F-ratio test.

If the variance of a population is known or can be estimated, the number of samples required to obtain a confidence interval of specified size can be calculated. Knowledge of the variance can come from other studies involving similar data or a small preliminary study.

ANALYSIS OF VARIANCE

Analysis of variance, which is often shortened to the acronym ANOVA, is a method used to compare several data sets. This is accomplished by comparing the degree of variability of measurements within individual sample sets to those among different sample sets to determine if their means are significantly different. The null hypothesis being tested is that all of the sample means are equal.

In biology and medicine, the different sample sets often represent different treatments (for example, does treatment with drug A produce better results than treatment with drug B or a placebo?). In geology, the samples

A mother with her triplets. The statistical chance of a woman having triplets without fertility treatments is about one in 9,000 births. SANDY FELSENTHAL/CORBIS.

might represent the sizes of fossils from different locations or the amount of gold in samples from several different rock outcrops. In political science, the samples might contain the ages of voters with different political tendencies (for example, are the average ages of liberal, moderate, and conservative voters significantly different?).

ANOVA assumes that the samples being compared are normally distributed (thus, like regression, it is a parametric procedure), that their variances are approximately equal, and that their samples are approximately the same size. Variances are calculated for each sample or treatment, and all of the samples are grouped together to calculate a total variance. ANOVA assumes that the total variance consists of two components: one resulting from random variance within each sample and the other resulting from variance among the different samples. The two variances are compared using an F-ratio test to determine whether the null hypothesis can be rejected at a specified level of significance. In the hypothetical case that all of the samples are identical, the variance among samples (and therefore the F-ratio) is zero. Thus, the null hypothesis would not be rejected. If the F-ratio is large, and depending on the sample sizes and desired level of significance, the null hypothesis may be rejected. As with all statistical tests, the F-ratio tests in ANOVA do not prove anything. They can only be used to reject or fail to reject the null hypothesis at a specified level of significance.

USING STATISTICS TO DECEIVE

The aphorism that there are "lies, damned lies, and statistics" is attributed to British statesman Benjamin Disraeli (1804–1881) and reflects the unfortunate fact that statistics can be accidentally or deliberately used to deceive just as easily as they can be used to illuminate and inform. Understanding how statistics can be accidentally or deliberately used to misrepresent data can help people to see through deceptive uses of statistics in real life.

Consider a group of four friends who graduated from the same college. Three of them earn $40,000 per year working as managers in a local factory, while the fourth earns $500,000 per year from his family's shrewd

Correlation or Causation?

Some of the most common examples of real life statistics are news stories describing the results of recently published medical or economic research. A newspaper article might give details of a study showing that men and women with college degrees tend to have higher incomes than those who have never attended college. A report on the evening news might explain that researchers have found a correlation between low test scores and excessive soft drink consumption among high school students. In both cases, variables are correlated but the studies do not necessarily prove that one causes the other to occur. In other words, correlation does not necessarily imply causation.

It is easy to think of reasons why people who obtain college degrees tend to make more money than those who do not. College degrees are required for many high paying jobs in science, engineering, law, medicine, and business. College graduates also know other college graduates who can help them to get good jobs and can take advantage of on-campus interviews. People who do not attend college, in contrast, are excluded from many high paying careers and may not have the same advantages as college students. This is not to say that there are no exceptions, because someone with a college degree may choose to take a low paying job for its intrinsic satisfaction. Social workers, teachers, or artists, for example, may have college degrees but earn less money than factory workers without degrees. Likewise, some multi-millionaires and even billionaires never completed college. What about the converse? Is it possible that high earnings cause people to become college graduates? In one sense, the answer is no. People usually attend college early in life, before they begin full-time careers, so it is unlikely that high earnings cause college attendance. It also seems unlikely that someone will make a sizable amount of money and, because of that, decide to attend college. It seems safe to conclude that, all other things being equal, college degrees are likely to cause higher earnings.

The other result, showing a correlation between soft drink consumption and low test scores, may be more difficult to explain. It is difficult to imagine that soft drink consumption alone causes a chemical or biological reaction that reduces intelligence and lowers test scores. But, there may be other factors to consider. It may be that students who like soft drinks place a higher priority on instant gratification than discipline, a quality that might also cause them to spend less time studying than students who consume few soft drinks. If that is the case, then both excessive soft drink consumption and low test scores are caused by another factor such as their parents' attitudes towards delayed gratification. If so, correlation would not reveal causation in this case.

investments in the stock market. What statistic best represents the income level of the four friends? The arithmetic mean is ($40,000 + $40,000 + $40,000 + $500,000)/4 = $155,000, but in this case the arithmetic mean is not an accurate reflection of the underlying bimodal population. If anything, the median income of $40,000 is more representative of most of the group even though it does not accurately reflect the highest salary. It is likewise strictly correct to state that the incomes of the four friends range from a minimum of $40,000 to a maximum $500,000, but that simple statistic does not convey the fact that most of the friends earn the minimum amount. It would therefore be true but misleading for a university recruiter to tell prospective students that a group of its graduates earns an average of $155,000 per year or that graduates of the university earn as much as $500,000 per year. A less deceptive statement that that the group earns between $40,000 and $500,000, and that three of them earn the minimum amount (or that the mode is $40,000). But, this still does not paint an accurate picture. An even less deceptive statement would also explain that while the highest earner is indeed a graduate of that college, his income is tied to his family's investments and not necessarily related to his college education.

There are several kinds of clues that can help determine if statistics are deceptive. The first is use of only maximum or minimum values to characterize a sample or population, to the exclusion of any other statistics. Parties involved in a dispute may emphasize that reported values are as high as or as low as a certain figure without giving the range, mean, median, or mode. Or, someone hoping to use statistics to prove a point may cite a mean without mentioning the median, mode, or range. Another potential source of deception is the use of biased or misrepresentative samples, which may produce sample statistics that are not at all representative of the underlying population. Reputable statisticians will always explain how their samples were chosen.

A Brief History of Discovery and Development

The history of statistics dates back to the first systematic collection of large amounts of data about human populations in the sixteenth century. This included weekly data about deaths in London and data about baptisms, marriages, and deaths in France. The first book about statistics, titled *Natural and Political Observations Upon the Bills of Mortality*, was written by the English mathematician John Graunt (1620–1674) in 1662. His motivation was practical: London had suffered from several outbreaks of plague, and Graunt analyzed weekly death statistics (bills of mortality) to look for early signs of new outbreaks. He also estimated the population of London. British astronomer Edmond Halley (1656–1742), best known for the comet that bears his name, wrote about birth and death rates for the German city of Breslaw (sometimes spelled Breslau, and now Wroclaw, Poland). His results were used by the English government to set the prices of annuities, which provided regular payments similar to a retirement fund, according to the age and sex of the person. The government had previously lost a considerable amount of money when it sold annuities to young people using rates based on average life expectancy during times of plague and war, and the annuity holders failed to die quickly enough. The French mathematician Abraham de Moivre (1667–1754) worked in London and was also interested in the statistics of death and annuities, publishing the book *The Doctrine of Chances* in 1714. He is known as the first person to write about the important properties of the normal distribution, and also for predicting the date of his death.

The dawn of the eighteenth century was marked by an explosion of inquiry about statistics in probability, including important books by Karl Friedrich Gauss (1777–1855) and Pierre Simon Laplace (1749–1827). The normal distribution is often known as the Gaussian distribution in deference to his work. The Statistical Society was established in London in 1834, and five years later the American Statistical Association was established in Boston. Much of the theory that stands behind modern statistics, though, was not discovered until the early twentieth century by notables such as Karl Pearson (1857–1936), A.N. Kolmogorov (1903–1987), R.A. Fisher (1890–1962), and Harold Hotelling (1895–1973), for whom numerous statistical methods and tests are named. One of the most unusual statisticians of the early twentieth century was William S. Gosset (1876–1937), who wrote under the pseudonym Student. He is best known for the t-test and t-distribution, which is commonly referred to as Student's t.

Real-life Applications

GEOSTATISTICS

Geostatistics is a specialized application of statistics to variables that are correlated in space, and is based on a concept known as the theory of regionalized variables. It has important applications in fields such as mining, petroleum exploration, hydrogeology, environmental remediation, ecology, geography, and epidemiology.

Traditional statistics is concerned with issues such as sample size and representativeness, but does not explicitly address the observation that many variables are spatially correlated. Spatial correlation means that samples taken in close proximity to each other are more likely to have similar values than those taken great distances apart. The variable being sampled might be the distribution of insect types or numbers across a landscape, the physical properties that characterize a good petroleum reservoir or aquifer, the occurrence of valuable minerals (such as gold or silver) in different parts of a mine, or even real estate prices in different parts of a city. Whatever their discipline, people who use geostatistics measure some variable at a limited number of points (for example, places where oil wells have already been drilled or the locations of homes that have been sold in the past few months) but need to calculate values at locations where they have no measurements. This process is known as interpolation, and geostatistics provides a set of tools that interpolate values based on the distribution of known values at different locations.

Central to the theory and application of statistics is the variogram, which is a graphical representation of spatial correlation. It depicts the variance among samples located different distances from each other, as opposed to the variance of an entire group of samples without regard to their locations. To calculate a variogram, samples are generally grouped or binned. For example, samples located between 0 and 100 meters from each other are put into one group, samples located between 101 and 200 meters from each other are put into a second group, and so forth. The distance between samples is known as the separation distance or lag. A variance is calculated for each group of samples, and the results are then plotted on a graph as a function of the separation distance. This is traditionally done using the semi-variance, which is one-half of the variance, rather than the variance itself.

If a variable is spatially correlated, the semi-variances will increase with separation distance and eventually reach a constant value known as a sill. The separation distance at which the sill is reached is known as the variogram range. The semi-variance will, in theory, decrease to zero when the separation distance is zero. This is because if one

repeatedly measured a value at the same location, the result should always be the same.

In real life applications, however, the result may differ if several samples are taken at the same location. If the values are chemical concentrations, for example, the differences may arise as a result of analytical errors or the inability to collect more than one sample (such as a scoop of soil) from exactly the same position. A non-zero semivariance at zero separation distance is known as a nugget or the nugget effect. This term dates back to the origin of geostatistics as a practical tool for mining engineers who needed to calculate the grade, or richness, of ore in order to determine the most efficient and economical way to run their mines. An unusually rich nugget or pod of ore might yield a very high grade, whereas rock or soil a very short distance away might have a much lower grade.

Once a variogram is developed, values can be interpolated using a process known as kriging, named after the South African mining engineer who invented the technique. Variograms can also be used as the basis for geostatistical simulation, which uses information about spatial variability to generate alternative realizations that are equally probable and poses the same statistical properties as the samples from which they are derived. A petroleum geologist might, for example, use geostatistical simulation to generate alternative realizations of an underground oil reservoir for which she has definite information from only a handful of wells. The exact nature of the oil reservoir between the existing wells is unknown, and geostatistical simulation provides a series of possibilities that can be used as input for computer models that determine how to most efficiently remove the oil.

QUALITY ASSURANCE

Statistics play a critical role in industrial quality assurance, and are often used to monitor the quality of products and determine whether problems are random occurrences or the result of systematic flaws that need to be corrected. All manufactured products will have some degree of variability. Components may be slightly shorter or longer than designed, not exactly the correct color, or prone to premature failure. Statistical process control can be used to monitor the variability of product quality by sampling components or finished products. If the results fall within pre-established limits (for example, as defined by a specified mean and variance), the process is said to be in control. If results fall outside of acceptable limits, the process is said to be out of control. Statistical quality analysts can also examine trends. If there is a gradually increasing number of unacceptable products, the underlying cause may be a piece of machinery that is gradually going out of adjustment or about to fail. Trends that fluctuate with time and appear to be correlated with factor shift changes may indicate human errors.

Six Sigma is an extension of statistical quality control that has evolved into a popular business philosophy. As it is used by many people, the term Six Sigma is nothing more than another way of saying that a process or procedure is nearly perfect or, among those who are slightly more mathematically inclined, that it produces no more than 3.4 failures per million opportunities. In the traditional manufacturing sense, each item produced on an assembly line is an opportunity to fail or succeed. In service-oriented fields such as retailing and health care, the opportunities might represent customer visits to a store or patient visits to a hospital.

The sigma in Six Sigma refers to the standard deviation of a normally distributed population, which is often represented in equations by the Greek letter sigma. The six has to do with the number of standard deviations required to achieve the desired standard of less than 3.4 failures per million opportunities.

Imagine that a bolt that is part of an airplane is designed to be exactly 10 centimeters long, but will still work if it is as short as 9.9 centimeters. Anything shorter than that will not fit and must be discarded. The owner of a machine shop hoping to supply bolts to the aircraft company collects samples of his product, carefully measures each bolt, and learns that the sample has a mean of 10 centimeters and a standard deviation of ±0.1 centimeter. If the owner collected a representative sample and bolt length that follows a normal distribution, then he can expect that 16% of the bolts will be too short. This is because 16% of a normal distribution is less than or equal to the mean minus one standard deviation, regardless of the size of the mean or the standard deviation. He can still provide bolts to the aircraft company, but would be forced to throw out 16% of his production to meet the standards. This amount of waste is inefficient and costs money, so the owner decides to adopt a Six Sigma policy.

To achieve Six Sigma, he must refine his bolt manufacturing process so that the standard deviation is small enough that only 3.4 out of each million bolts produced (or 0.00034%) are less than 9.9 centimeters. For a normal distribution, 0.00034% of the population is less than the mean minus 4.5 standard deviations, or 4.5 sigma. The average length of bolts produced in the machine shop, though, varies over time. This might be the result of seasonal temperature fluctuations (metal expands and contracts as its temperature changes), small variations in the composition of the metal used to make the bolts, or a host of other factors. Pioneering studies of electronics manufacturing

Cellular Telephones and Political Polls

Political pollsters have long relied on telephone surveys to sample public opinion on matters ranging from presidential elections to advertising effectiveness. As long as virtually everyone has a telephone, the population of a city, region, or nation can be sampled by randomly selecting telephone numbers and calling those people. Even people with unlisted telephone numbers are fair game because pollsters can use computers to generate and dial telephone numbers. Although there are some people without any telephone service, they generally represent less than 5% of the population.

The explosive growth of cellular telephone use, and particularly the increasing number of people who use only cellular telephones and do not have land line telephones, became an issue in the 2004 United States presidential election. During the months leading up to the election, some experts believed that a disproportionate number of people who used only cell phones were young voters. This presented a problem because political pollsters do not call cellular telephones. Federal law makes it illegal to use automated dialing machines to reach cellular telephones, and some state laws prohibit unsolicited calls to numbers at which the recipient will have to pay for the call (which includes most cellular telephones). If each voter is equally likely to have only a cellular telephone, then survey results will not be affected. If certain segments of the population, however, are more likely than others to be inaccessible to pollsters then the reliability of their polls decreases because their samples will be biased. The influence, if any, of young cellular-only voters on pre-election polls for the 2004 presidential election was never conclusively determined. The potential for poll bias as growing numbers of people abandon their traditional land line telephones for cellular phones, however, promises to be an important consideration in future elections.

processes showed that the mean value must be 6, not 4.5, standard deviations away from the acceptable limit in order to ensure no more than 3.4 defects per million products. In others words, an additional increment of 1.5 standard deviations is added to account for the fluctuations. Hence the association of the name Six Sigma with a defect rate of 3.4 pieces per million. In terms of the bolt manufacturer, this means that he must improve his manufacturing process to the point where the standard deviation of bolt lengths is $(10.0 - 9.9)/6 = 0.017$ centimeters.

PUBLIC OPINION POLLS

Public opinion polls, particularly political polls during major election years, are another real life application of statistics in which samples consisting of a few hundred people are used to predict the behavior or sentiments of millions. Careful selection of a representative sample allows pollsters to reliably forecast outcomes ranging from consumer product demand to election outcomes.

Modern public opinion polling starts with carefully selected questions designed to elicit specific opinions. For example, asking a voter whether she likes Candidate A may elicit a different response than asking the same voter if she dislikes Candidate B, even if Candidate A and Candidate B are the only choices. Interviewers are trained to ask questions in a neutral, rather than suggestive or leading, manner. The selection of people to be interviewed, known as sampling, begins with the generation of random telephone numbers. Known business telephone numbers and cellular telephone numbers are removed from the list, and random number generation ensures that every residential telephone number has an equal probability of being called even if it is not listed in the telephone directory. In a national poll, the list of telephone numbers is then sorted by state and county and the number of telephone numbers called for each state or county is proportional to its population. Because there may be more than one eligible respondent in each residence, interviewers may ask to speak to the person who has had the most recent birthday. Women are more likely than men to provide complete and usable responses, so interviewers ask to speak to male household members more often than female household members to account for that bias.

The number of people interviewed is estimated using a standard formula based on the normal distribution. The formula predicts that the uncertainty of results (often referred to as the margin of error) for a random sample of 500 people, which is a common size for a nationwide political poll in the United States, is ±4.4%. The uncertainty is inversely proportional to the square root of the sample size, so increasing the sample size to 5000 (a factor of 10) decreases the margin of error to ±1.3% (a factor of 3.4). Decreasing the sample size to 50

would increase the margin of error to ±14%. Thus, the often used sample size of 500 represents a compromise that provides relatively reliable results for a reasonable expenditure of time and money.

Once the required number of responses have been obtained, the results are broken down into groups according to the age, race, sex, and education of the respondent. The results for each group are weighted according to census results in order to arrive at a final result that is representative of the population as a whole. For example, if 30- to 40-year-old Asian males who graduated from college comprise 2% of the population but represent 3% of the poll respondents, then the results are adjusted downward so that they do not unduly influence the outcome.

Perhaps the most difficult political polling problem is the identification of so-called likely voters. Pollsters will ask respondents if they are likely to vote in an upcoming election, but there is no guarantee that the respondent will follow through. Unexpected bad weather, in particular, can reduce the number of voters and skew results if different parts of the country are affected. Good weather in states with many conservative voters may compound bad weather in states with many liberal voters, or vice versa. Unexpected mobilization of large blocs of voters with vested political interests, for example religious or labor groups, may also invalidate pre-election polls. Thus, the political pollster is faced with the problem of trying to sample a population that will not exist until election day.

Potential Applications

The potential applications of statistics in real life will increase as society continues to rely on technological solutions to social, environmental, and medical problems. Optimization methods based on statistics are becoming increasingly more important as airlines strive to become more competitive. Advance knowledge of the likely weight of passengers and their luggage, or the number of passengers who are likely to miss their flights, can help an airline to utilize its resources in the most effective manner possible. High tech manufacturing calls for rigorous quality assurance procedures to ensure that expensive and complicated electronic components don't fail, especially those used in situations where failure may have life-threatening consequences. The explosive growth of the Internet during the 1990s led to the creation of a new field known as data mining, which involves the statistical analysis of extremely large data sets containing many millions of records, that will no doubt continue to grow as the prevalence of electronic commerce increases.

Where to Learn More

Books

Best, Joel. *Damned Lies and Statistics: Untangling Numbers from the Media, Politicians, and Activists.* Berkely: University of California Press, 2001.

Graham, Alan. *Teach Yourself Statistics (2nd edition)* Chicago: McGraw-Hill, 2003.

Huff, Darrell. *How to Lie With Statistics (reissue).* New York: W.W. Norton, 1993.

Periodicals

Langer, Gary. "About Response Rates," *Public Perspective.* vol. 14 no. 3 (2003): 16–18.

Web sites

ABC News. "ABC News' Polling Methodology and Standards." 2005. <http://abcnews.go.com/US/PollVault/story?id=145373&page=1> (April 9, 2005).

UCLA Department of Statistics. "History of Statistics." August 16, 2002. <http://www.stat.ucla.edu/history/> (April 9, 2005).

Subtraction

Overview

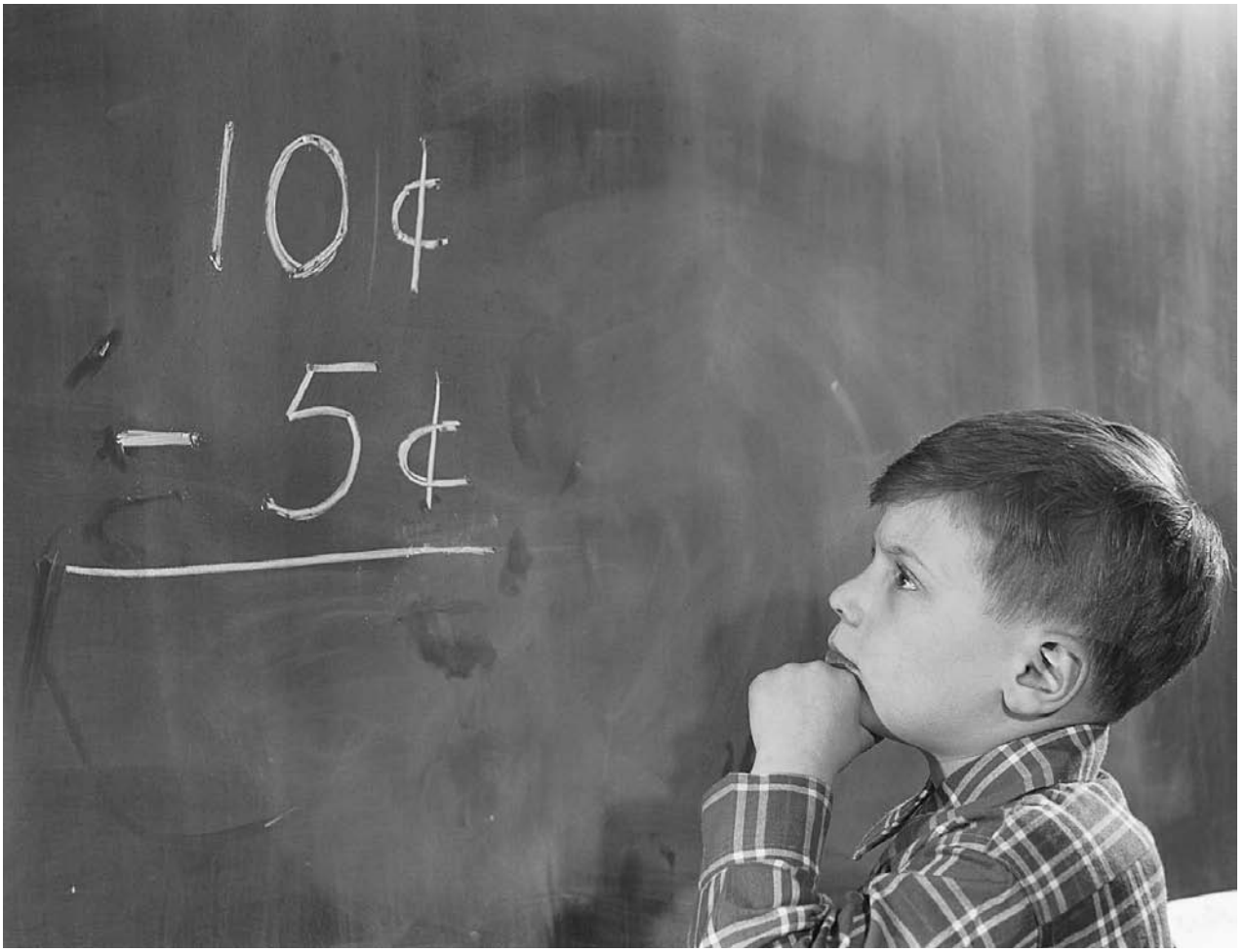
Subtraction is the inverse operation of addition. It provides a method for determining the difference between two numbers; put another way, it is the process of taking one number from another to determine the amount that remains. While the basics of this fundamental process are taught at the preschool level, subtraction provides a foundation for many aspects of higher mathematics, as well as a conceptual basis for some cutting-edge methods of developing new technology. In addition, subtraction provides answers to a wide array of practical daily questions in areas ranging from personal finance to athletics to making sure one gets enough sleep to remain healthy.

Fundamental Mathematical Concepts and Terms

A subtraction equation consists of three parts. The solution or answer to a subtraction equation is called the difference. While this term is commonly known, the other two elements of a subtraction equation also have labels, albeit far less well-known ones. The starting value in a subtraction equation is called the minuend, while the second term is called the subtrahend. Thus, a subtraction equation is formally labeled: minuend − subtrahend = difference. Simple two-place subtraction problems can be solved by subtracting each column individually, beginning at the right and working progressively left. The equation 49 − 21 is solved by evaluating 9 − 1 for the right value and 4 − 2 for the left value to produce a final answer of 28.

Complications in this simple process arise when borrowing and carrying become necessary, as in the equation 41 − 28. Because 8 cannot be directly subtracted from 1, it becomes necessary to borrow ten from the next place, in this case the value 4. This operation is made possible by applying the distributive property of mathematics, that describes how values can be distributed in multiple ways and that in this example insures that the value 41 is equivalent to the expression 30 + 11. Following this operation, solving this equation is simply a matter of subtracting 8 from 11 and 2 from 3 using the same column by column approach demonstrated in the initial example. Subtraction equations using large values may require multiple instances of borrowing in order to produce a solution, though the method used to solve these equations is identical to that used for simpler equations.

A second complication arises when subtraction involves negative numbers. While the physical world does not contain negative quantities of any physical object, some measurement systems include negative values, the



While the basics of this fundamental process are taught beginning at the preschool level on up, subtraction provides a foundation for many aspects of higher mathematics, as well as a conceptual basis for some cutting-edge methods of developing new technology. WILLIAM GOTTLIEB/CORBIS.

most common of these being the modern system of temperature measurement. Whether dealing with Fahrenheit or Celsius, both systems measure temperature with values gradually falling to a value of 0 long before temperatures stop decreasing; in both systems, the temperatures reach zero and simply begin again, this time with the number values labeled negative and decreasing as the temperature cools, such that −10 degrees is colder than 10 degrees.

Now suppose that we wish to find the difference between a day's high and low temperatures, or the temperature range for that day (also called the diurnal temperature). If the high and low temperatures are both positive, this is accomplished by simply subtracting the low temperature from the high temperature to find the difference. However, if the low value happens to be negative, this process must be handled differently. In order to subtract a negative number, we simply add the absolute value of that number; if we wish to subtract −14, we accomplish this by adding 14. Applying this convention to a day where the high is 40 and the low is −9, we solve this equation: $40 - (-9)$, that we convert to $40 + 9 = 49$, the difference in the two measured temperatures and the temperature range for the day. This same process can be used for any temperature system that does not have an absolute 0 point, as well as in any other type of measure that uses both positive and negative values. Among modern temperature scales, the only one that does not require this type of adjustment is the Kelvin temperature scale, in which 0 represents the coldest any object can ever become, and the point at which molecules have a minimum of molecular motion (many texts incorrectly state that at absolute zero motion ceases. However, this is incorrect because there is still vibratory motion). For comparison, 32 degrees Fahrenheit (32°F) equals 0 degrees Celsius (0°C), and equals 273 Kelvin (273 K).

Because carrying is frequently required to resolve subtraction equations, most people find subtraction harder to

perform than addition. For this reason, a different type of borrowing and carrying is sometimes employed to simplify mental subtraction. In an equation such as 41 − 29, the first step requires borrowing ten and adding it to the 1, the step at which most mistakes are made, and where a simple shortcut can help avoid errors. This shortcut is based on the fact that the simplest number to subtract from any value is 0, and this shortcut takes advantage of this fact. To apply this shortcut to the equation 41 − 29, we simply change the 29 to 30 by adding one. Then, we can easily evaluate the new equation, 41 − 30, to get 11, to which we add back the one extra that we subtracted to reach the correct total of 12. This process can be quickly learned, and with practice becomes routine, helping improve the accuracy of mental arithmetic.

A Brief History of Discovery and Development

Subtraction has been used for millennia, initially being calculated with counting sticks, stones, or other items, and later using early tools such as the counting table and the abacus. However, the written notation for subtraction, the modern minus symbol, came into use much more recently. In England during the 1400s, the dash as a minus symbol was first used to mark barrels that were under-filled, signifying that the marked barrels had missing or inadequate contents. By the 1500s, this notation had migrated from barrels into mathematical notation as the accepted symbol for subtraction, and has remained in use ever since.

The modern method of solving subtraction problems can be traced as far back as the 1200s, when this method was originally called decomposition; not until the 1600s did the term "borrowing" come into use. Two other subtraction methods were also taught well into the twentieth century, though these are largely forgotten today. One fairly intense debate arose during the early 1900s, dealing with the proper notation for subtraction. While students today are taught to cross out values and write in new ones above them as part of the borrowing process, this practice did not appear widely in American textbooks prior to the middle of the twentieth century. Before this adoption, an ongoing debate raged over the use of these hash marks, or crutches as they were originally called. Critics argued that subtraction should be accomplished without the use of this pejoratively labeled aid; one 1934 math text went so far as to give examples of equations performed both with and without "crutches," labeling the version without crutches the preferred method and noting that teachers should not allow students to use crutches when solving problems. Advocates of crutches, many of them school teachers, based their argument on simple utility, countering that the use of crutches aided students in calculating correct results with fewer errors. A 1930s study published by researcher William Brownell offered strong evidence that the teachers were right, and that using crutches or other notations to keep track of borrowing did reduce errors in subtraction. Almost immediately following this study, textbooks teaching the crutch notation method of subtraction became the norm, and this technique continues to be used today.

Real-life Applications

SUBTRACTION IN FINANCIAL CALCULATIONS

Profit is the amount earned from a business transaction, and can be found using subtraction. In the simplest form, profit is determined by subtracting cost from selling price; for the up-and-coming lemonade merchant who takes in $6.75 from her customers after spending $2.25 on lemonade mix, cups, and ice, a simple profit calculation of $6.75 − $2.25 reveals a positive outcome or profit of $4.50. However, profit calculations are rarely this simple, and in many cases, unplanned costs can subtract significant amounts from the final profit earned.

Consider a beginning entrepreneur trying to make a start on E-bay. This young businessman purchases the latest Tony Hawk PlayStation game at a garage sale for $14.00. Because he already owns a copy of this game, he is eager to sell it on E-bay for a quick profit. He lists it on the auction site with free shipping and a "Buy-it-now" price of $19.95 that he calculates will give him a quick $5.00 profit after paying his expenses. The game sells quickly, the seller ships it to the buyer, and then sits down to calculate his profits.

The beginning point of this calculation is the amount of income received, often called gross income, that in this case is $19.95. From this starting value, the seller must subtract all his expenses to find his actual profit, sometimes referred to as net income. He begins with his cost for the game, that was $14.00; 19.95 − 14.00 = 5.95. From this value, he then subtracts his other costs, such as postage of $1.45; 5.95 − 1.45 = 4.50. The seller was surprised to find that the padded envelope he needed was more expensive than he expected, at 75 cents; 4.50 − .75 = 3.75. Other fees also must be subtracted, and while most of these are small, they begin to accumulate. E-bay fees including a listing fee, "Buy-it-now fee," additional photo fee, and final sale fee totaled 1.75; 3.75 − 1.75 = 2.00. The final surprise for the young businessman comes

when he receives his electronic billing statement and learns that the service charged him 3% of the total sale price of $19.95, or 60 cents; 2.00 − .60 = 1.40. The final profit left after subtracting all expenses is $1.40, far less than he had hoped. What appeared to be a fairly profitable business transaction turned out to be a near-loss when all the relevant expenses were correctly subtracted.

TAX DEDUCTIONS

One of the more enjoyable uses of subtraction involves the use of tax deductions. Throughout history, most taxpayers around the world have complained that taxes are too high. In the American federal tax system, several items may be subtracted from total income before taxes are calculated, and in many cases, the net tax savings from these items can be thousands of dollars.

The standard U.S. Federal Income Tax form is called Form 1040. On the first page of this form, taxpayers enter the total amount of their earnings for the year. However, before paying taxes, numerous items are subtracted, reducing the taxable income as well as the actual income tax paid. For instance, taxpayers are allowed to take a personal exemption for each family member; for tax year 2004, this exemption is $3,100, meaning that a family of four can subtract $3,100 four times, for a total reduction in taxable income of $12,400. Contributions to an Individual Retirement Account are often deductible up to a maximum limit (e.g., $3,000 per person), and self-employed individuals (those who don't work for a company) can deduct their costs of health insurance from their taxable income. In many cases, students can deduct tuition and textbook costs up to the maximum allowed limit as well. Finally, expenses such as mortgage interest on a home loan can be deducted prior to calculating the actual tax bill.

Only after all these items and others are deducted, or subtracted from gross income, is a final value reached. This value, called taxable income, is the actual amount on which federal taxes are calculated. Because so many items can be subtracted before calculating taxes, a typical family of four might easily reduce its taxable income by $20,000 or more by following the tax instructions carefully. Because the tax system is designed with these subtractions as an expected part of the process, failing to claim these deductions is equivalent to voluntarily paying more income taxes than required, something very few taxpayers have any interest in doing. Modern tax software has made the previously tedious process of tax filing far simpler and more accurate.

Along with electronic tax filing, some tax services offer to give filers their tax refund immediately, in the form of a refund anticipation loan or RAL. RALs are offered to tax filers who don't want to wait for their tax refund to arrive. While RALs may be a useful tool for situations in which money is needed immediately, an RAL can significantly reduce the amount of the final refund. For example, a consumer expecting a tax refund who requests an RAL would typically have to subtract several fees, including an application fee that averages about $30, and a loan fee that can range from $30 to more than $100. For 2005, a refund of $2,050 incurred an average fee of $100, which reduces the total refund to $1,950. While this reduction seems small, it represents a 5% fee for borrowing this money until the actual refund arrives from the IRS. Because the average refund is now deposited in less than two weeks, this loan equates to an annual percentage rate of roughly 187%. In 2003, 11% of taxpayers took RALs, costing themselves more than $1 billion in fees for these short-term loans that many consumer advocates criticize as an unreasonable effort to charge taxpayers interest on their own money.

Rebates are a popular method of selling an item for less than its original price in order to attract buyers. Rebates come in several forms. Most new cars today are sold with a manufacturer's rebate, meaning that the sticker price on the window of the car is automatically reduced by subtracting the rebate amount. This rebate is in addition to the normal amount subtracted from the sticker price by most car dealers. Automobile rebates are paid automatically to any buyer, and are given at the time of purchase. Information on actual dealer costs and available rebates can be found at numerous online car buying sites.

Another popular form of rebate is the mail-in rebate. These rebates are frequently offered on electronic equipment and other high-priced items, particularly in the case of older merchandise that manufacturers wish to clear out of inventory. A mail-in rebate is not paid at the time of purchase; instead, the purchaser is required to complete one or more rebate forms and mail these forms, along with specific pieces of documentation, to a processing center. If the documents are submitted correctly and prior to the offer's deadline, a check is normally mailed to the buyer within a period of four to six weeks.

Why are mail-in rebates so popular with manufacturers, and why do companies use rebates instead of simply reducing the price of the products? Consumers behave in predictable ways, and most rebate programs save manufacturers money due to a phenomena researchers call slippage, in that many customers never redeem their rebates. Estimates vary on just how high slippage rates are, and the rate is influenced by factors such as the size of the rebate, the length of time allotted to redeem it, and the difficulty of complying with the program rules. However, on

average, rebate redemption rates for small items can be as low as 2%, while for larger rebates in the $50 to $100 range, redemption levels typically hover around 50%. The benefit of rebates to the manufacturer are obvious: they can advertise a much lower price, knowing that half or fewer of the buyers will get this lower price, while the rest will pay the full, unrebated amount. Rebates can be a wonderful bargain for those who follow through on them. However, for many buyers, the promised reduction in price is never realized due their own unwillingness to follow through on the process.

While most highways can be driven free of charge, toll roads require a driver to pay for the privilege. While using a toll road has traditionally meant stopping to throw a handful of coins into a basket or waiting for an attendant to make change, many toll roads now provide the option to pay electronically without stopping. These systems, with names such as Pike Pass in Oklahoma and FasTrak in California, allow a user to purchase a small electronic unit to mount in her vehicle; this unit can then be filled by paying in advance and then used like a debit card while driving. To use the automated systems, drivers typically change into a specific lane that is equipped with sensors to read data from the user's transmitter. Using this identification data, the system automatically subtracts the proper toll amount from the user's account; in many cases, the system automatically sends a reminder e-mail or letter when the balance drops below a set limit. Drivers using these systems not only avoid the hassle of carrying correct change with them and waiting in line to pay, some states also give them a reduced toll rate for using the automatic system. In addition to saving 5–10% on their tolls, drivers in Oklahoma also enjoy the pleasure of paying the toll while never dropping below the 75 mile per hour posted speed limit on the state's tollways.

A countdown clock on the Eiffel Tower in Paris marking the last 100 days before the year 2000. Countdown clocks use simple subtraction to countdown to zero. AP/WIDE WORLD PHOTOS. REPRODUCED BY PERMISSION.

SUBTRACTION IN ENTERTAINMENT AND RECREATION

One of the more entertaining uses of subtraction is a process known as a countdown, in that a large starting value is gradually reduced by one until it finally reaches zero. Countdowns are used in a variety of settings in that people need to know in advance when a particular event will happen. Countdowns are perhaps best known for their use in space exploration, where an enormous clock traditionally ticks off the final seconds until liftoff. While this process provides dramatic footage for television news, the use of countdowns, which typically start several days before launch, is actually a method of insuring that the complex series of events required for a successful launch are completed on time and in the proper sequence. Space launch countdowns normally include several planned holds, during which the countdown clock stops for a set period of time while various checks are made.

Countdowns are also used for recreational purposes. Each year, millions of people across the globe eagerly count down the final seconds until the arrival of a new year, celebrating its arrival with cheers, hugs, and toasts. Hockey players, banished to the penalty box for rule violations, sit and impatiently wait for their penalty time to count down to zero so they can re-enter the game. Top ten lists, including television host David Letterman's long-running version, are often used to poke fun or entertain by leading the audience gradually down from ten to one, and weekly top 20 countdowns guide music fans gradually to number one, the week's top song.

Golf Handicaps While most sports force players to compete head-to-head without any adjustment to the score, a few events attempt to level the playing-field by adjusting

player totals. Golf is one of the more popular sports in which subtraction is used to allow players of differing skill levels to compete on an equal basis. Using a system known as handicapping, a golfer's handicap index is assigned based on a series of ten recent rounds he has played. Using these game scores, a difficulty rating for the courses on which they were played, and a complex formula, an authorized golf club can issue an official handicap index to a player. Using this index, each player can then calculate his handicap for a particular course, meaning he is given strokes and can subtract a specific number of shots from his score. Using this system, a golfer who normally scores 76 and a golfer who normally scores 94 can compete fairly on the same course. By subtracting the proper number from each score, each golfer is able to arrive at an adjusted score and compare how well or how poorly he played that particular course that day.

Track and Field One measure of an athlete's performance is his vertical jump. Vertical jump is not a measure of how high an athlete can leap in absolute terms, because this result is strongly influenced by an athlete's height and arm-length; rather, vertical jump is a measure of how high an athlete can propel himself from a standing start, relative to his standing height; for this reason, it provides a better measure of absolute jumping ability than a simple measure of how high a leaping athlete can reach.

Vertical jump is calculated using subtraction. First, an athlete's standing reach is measured by having him stand flat-footed and reach as high as possible with one arm. Then, the athlete's jump reach is measured by having him stand and jump straight up without taking a step. True vertical jump is calculated using the following equation: Jump Reach − Standing Jump = Vertical Jump. For reference, professional basketball players typically have a standing vertical jump of 28–34 inches, meaning their final reach height is almost three feet higher than their standing reach. Jumping, like most other athletic skills, can be improved with training. Because of the explosive nature of jumping, performance is often improved using both strength-building and power-enhancing forms of exercise.

Pop Culture

Each December, millions of people around the world plan for a new year by making one or more New Year's resolutions. While many of these resolutions focus on addition, such as making more money, spending more time with family, or playing more golf, the two most popular resolutions for 2005 both involved subtraction. The second most popular resolution in 2005 was to lower payments by reducing personal debt. The most popular resolution has stood atop the list for some time, and will probably remain there: more people chose subtracting pounds, or losing weight, than any other New Year's resolution for 2005.

Weight Loss and Dieting

Because losing weight is such a popular goal, one might assume that many people are reaching this goal and losing weight. In truth, the popularity of the goal is probably tied to the increasing incidence of obesity; as of 2000, approximately two-thirds of United States adults were defined as overweight or obese, and predictions suggest that this number will continue to rise. Most of the hundreds of methods of subtracting pounds involve subtracting from what is being eaten. Some diets reduce intake of fats while others restrict intake of carbohydrates. While debate continues to rage on which plans work best (and that do not work at all), one piece of advice seems to make sense: reducing the amount of food on one's plate helps many people eat less. This simple subtraction can provide a solid starting point for any weight-loss plan, and has been shown to lead to weight loss even without any other behavioral changes.

Sleep Management

Before the invention of the electric light bulb, Americans slept an average of nine hours per night; today, the average is one to two hours less. While doctors and sleep experts recommend that teenagers get 8.5–9 hours of sleep each night, the average teenager in America gets far less. Sleep experts say that each person has a set need for sleep each night, and that each hour of missed sleep adds up to create a sleep deficit. This deficit describes how far in debt a person is in terms of sleep and represents needed sleep hours that have been subtracted and applied to other activities. While being a few hours overdrawn on sleep is not an immediate danger and can usually be made up over a weekend of sleeping late, the long-term impact of inadequate sleep can be serious. As the sleep deficit grows, a variety of negative physiological outcomes become more likely, including obesity, high blood pressure, reduced productivity at work, poor mood, and increased an likelihood of accidents at home, at work, and while driving. While sleep time can be subtracted over the short-term without major impact, the sleep account must eventually be rebalanced by adding additional hours of sleep to the account.

Subtraction in Politics and Industry

DOOMSDAY CLOCK

One famous countdown clock has been ticking for more than half a century, though this clock has actually moved only a few minutes during that time, and has occasionally run backwards. In June of 1947, the Bulletin of the Atomic Scientists, an academic journal dealing with atomic power and physics, placed on its cover a clock, with the hands showing seven minutes until midnight. In a lengthy editorial inside, the journal described this so-called Doomsday Clock, in which midnight signaled the destruction of mankind by atomic weapons. The Doomsday Clock stirred a great deal of discussion with its appearance during the earliest years of the atomic age.

In the years since 1947, the Clock has made many appearances on the journal's cover, with the minute hand moving either forward or backward depending on the state of world events. In 1949, after the Soviet Union detonated its first atomic weapon, the clock advanced four minutes, displaying three minutes before midnight. Four years later, following the test detonations of thermonuclear devices in both the Eastern and Western hemispheres, the hands advanced again, reaching two minutes until midnight. During the following years, events including new arms treaties and the rekindling of old conflicts nudged the minute hand repeatedly backward and forward. The signing of the Strategic Arms Reduction Treaty (START) in 1991 moved the clock to seventeen minutes till midnight, its earliest point ever. At its last appearance in 2002, the clock stood once again at seven minutes till midnight.

Engineering Design

As popular as weight loss goals are for individuals, subtracting pounds or ounces can also become a major goal in industry. During the design phase of the Apollo moon missions, NASA became concerned that the Lunar Module, the ship that would carry two astronauts on the final leg of the trip to the moon's surface, was significantly overweight. Major redesigns began, and, by reducing the size of the observation window, cutting the thickness of the craft's skin, and making other changes, the craft's weight was significantly reduced. However, in order to reach the specified weight target, Grumman, the craft's builder, resorted to extraordinary measures, at one point actually paying company engineers a bonus for each ounce they were able to shave off the craft's weight. The efforts of these professionals were successful, and the lunar module performed as designed.

Weight reduction is also a priority in the automobile industry. In order to meet fuel economy goals, most automobile manufacturers have made significant changes to their designs in order to subtract from the vehicle's total weight. In many cases, steel has been replaced with aluminum, which is more expensive, but far lighter; in other cases, plastics or lightweight carbon composites have been introduced in order to reduce weight. One extreme example of this type of engineering weight loss involves a revolutionary car, General Motors' EV1, the first totally electric production car. Introduced in 1996, the EV1 was also faced with extraordinarily tight weight limits in order to reach its target mass of under 3,000 pounds (1,360 kg). Toward this end, GM engineers adopted a variety of changes to subtract weight from the vehicle. Among the solutions was the decision to use aluminum for the frame and wheels, shaving more than 300 pounds (136 kg) off the weight of traditional steel parts, and the choice of a non-traditional material, magnesium, for the steering wheel and seat-backs. While the EV1 was not a commercial success, GM's experience in cutting weight during its development has led to applications in other vehicles. According to one calculation, an automaker can subtract $4.00 from a car's cost for each pound of weight it manages to remove from the design.

Potential Applications

While the basic process of subtraction itself offers few potential breakthroughs, the concept of removing items from a collection in order to reach an objective remains useful, and one early application of this principle is already producing impressive breakthroughs. Evolutionary design is a technique that allows computers to consider millions or billions of possible solutions to a complex problem to arrive at an optimal solution. In many ways, this process is similar to the concept of natural selection, in which the stronger predator survives to reproduce and pass his genes on to succeeding generations while the weaker predator is eliminated from the gene pool.

Antenna Design

The field of antenna design is unfamiliar to most people. However, the ability to design lightweight, efficient antennas is critical to the space program and other industries. One challenge in this endeavor has been that

antenna design requires a deep understanding of the field, limiting this work to a relative handful of experts. A second limitation is that even these experts are not always certain how to improve the design of a specific antenna. Evolutionary design accepts that the present understanding of how to improve antennas is limited; this process instead simply creates and evaluates so many different choices that it is likely to produce a useful one.

The evolutionary design process begins with a researcher creating a group of antennas with different combinations of shapes and sizes, that are then mathematically described for the software. Next, the software applies random mutations to these beginning antennas, such as lengthening some and giving others more or fewer arms. After that, the resulting antennas are tested for performance. Using the results of this testing, the more effective models are kept, while the poorest performers are replaced with new samples similar to the good performers. Then, the process of mutating the designs, testing the resulting models, and retaining the best versions is repeated. After this process of evolutionary improvement has occurred for thousands of generations, a single model eventually emerges that offers the best possible combination of performance traits.

In the case of this small, one-inch square antenna designed for satellite use, more than ten hours of supercomputer time was required to assess millions of possible configurations; by comparison, an expert antenna designer would have needed twelve years working full-time to process the first 100,000 designs. Further, given the strange appearance of the antenna, which resembles little more than a collection of strangely bent paper clips, it seems doubtful that a human designer would ever have proposed such a configuration. The secret to this unique design process lies in a radically advanced form of subtraction that allows removal of the every design except the very best ones, allowing those designs to be further enhanced. Future uses of this technique are anticipated in producing such developments as computer chips that can heal themselves in the case of malfunction, and improved components for implantable medical devices.

Where to Learn More

Books

Brownell, W.A. *Learning as Reorganization: An Experimental Study in Third-grade Arithmetic*. Durham, NC: Duke University Press, 1939.

Periodicals

Ross, Susan, and Mary Pratt-Cotter. "From the Archives: An Historical Perspective" *The Mathematics Educator*. (2000): 10 (2).

Shaw, Mary, Richard Mitchell, and Danny Dorling. "Time for a smoke? One cigarette reduces your life by 11 minutes." *British Medical Journal*. (2000): 320 (53).

Web sites

About Golf. Golf handicaps, an overview. <http://golf.about.com/cs/handicapping/a/handicapsummary.htm> (March 19, 2005).

Bulletin of the Atomic Scientists. Doomsday Clock. <http://www.thebulletin.org/doomsday_clock/timeline.htm> (March 17, 2005).

Centers for Disease Control. Obesity Trends Among U.S. Adults Between 1985 and 2003. <http://www.cdc.gov/nccdphp/dnpa/obesity/trend/maps/obesity_trends_2003.pdf> (March 19, 2005).

Federation of American Scientists. Strategic Arms Reduction Treaty. <http://www.fas.org/nuke/control/start1> (March 19, 2005).

Internal Revenue Service. Form 1040. <http://www.irs.gov/pub/irs-pdf/f1040.pdf> (March 17,2005).

The Math Lab. Subtraction in your head! An algebraic method for eliminating borrowing. <http://www.themathlab.com/Pre-Algebra/basics/subtract.htm> (March 19, 2005).

National Air and Space Museum Oral History Project. Interviewee: James E. Webb, November 4, 1985. <http://www.nasm.si.edu/research/dsh/TRANSCPT/WEBB9.HTM> (March 19, 2005).

National Sleep Foundation. Myths and Facts About Sleep. <http://www.sleepfoundation.org/NSAW/pk_myths.cfm> (March 19, 2005).

Spaceref.com. Press Release: NASA Evolutionary Software Automatically Designs Antenna. <http://www.spaceref.com/news/viewpr.html?pid=14394> (March 19, 2005).

U.S. Department of Energy; EV America. General Motors EV1 Specifications. <http://avt.inel.gov/pdf/fsev/eva/genmot.pdf> (March 18, 2005).

Symmetry

Overview

Objects that have parts that correspond on opposite sides of a dividing line are said to have symmetry.

Fundamental Mathematical Concepts and Terms

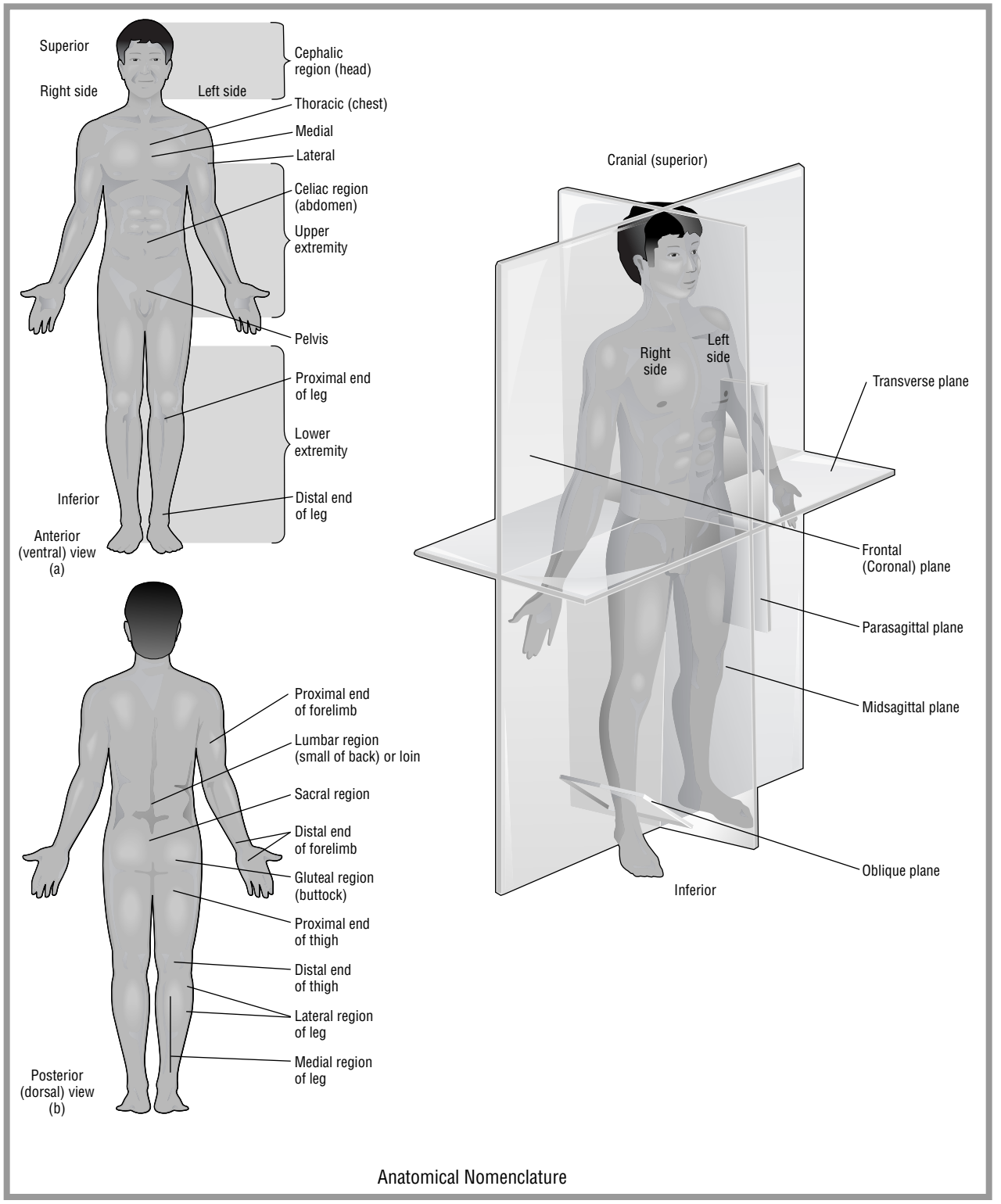
If a spatial operation can be applied to a shape that leaves the shape unchanged, the object has a symmetry. There are three fundamental symmetries: translational symmetry, rotational symmetry, and reflection symmetry.

An example of translational symmetry can be seen in lengths of rope or in the patterns on animals. If the rope is closely inspected, a braided pattern can be seen. By moving along the rope a bit further, the same pattern is seen again; thus the rope has translational symmetry. This pattern is very important for climbers, if the braided pattern is distorted in any way the force will no longer be evenly distributed along its length and it can break at this point under load. For this reason, climbing ropes will often have brightly colored patterns in their braiding to help the climber spot any deviations from this symmetry.

Imagine a sunflower that is the object of an operation, and the operation can be applied to its rotation around the center of the flower. If it is rotated so that the petals line up again so that it will look the same as before, the sunflower pattern is said to be "symmetric under rotation." Symmetries are probably the easiest patterns in nature for us to see and also the most common. The reason that nature has used symmetry in such abundance is that it allows complex objects to be constructed from simpler shapes, greatly reducing the amount of information that needs to be stored and processed to build the object.

Your whole body has reflection symmetry along the center. This symmetry can be seen if you stand by a reflective shop window, or large mirror, so that one half of your body is hidden from view and the other half is reflected. To an observer it looks as if you are whole because humans have a biological symmetry (often distorted or fused in the case of internal organs such as the heart) that roughly corresponds to an imaginary plane through the sagittal suture of the skull that divides the body onto left right planes.

Other symmetries can be built by repeated application of these basic symmetries, for example, the teeth of a zipper have a symmetry made by reflection and translation. This symmetry is called glide-reflection.



A plane through the sagittal suture establishes a plane of left and right symmetry for the human body. ILLUSTRATION BY ARGOSY. THE GALE GROUP.

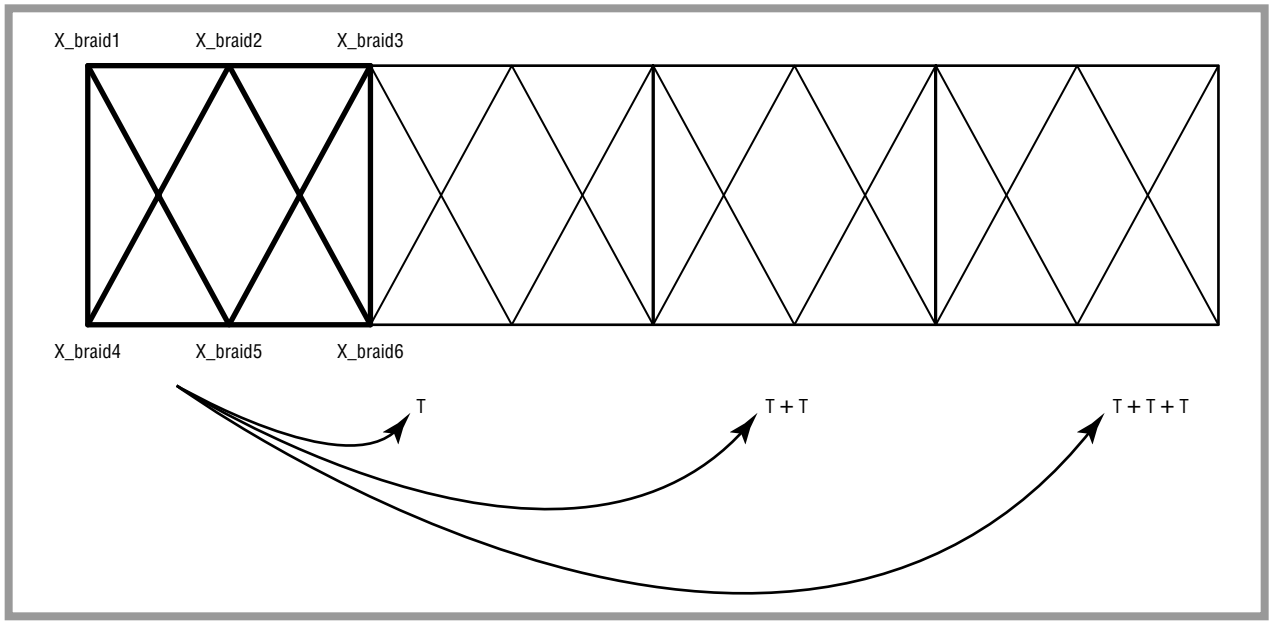


Figure 1.

EXPLORING SYMMETRIES

To understand the nature of translation, rotation, and reflection symmetry, one must first define how these operations act on an object. If an object is defined by a set of points, an operation can be defined by its action on these points.

Let us start with translation, the basic braided pattern of a rope can be recorded by a number of points which can be grouped together into a set called X_braid. As a simple braiding, imagine the rope has a repeating pattern made from two crosses inside by a box. This pattern can be represented by points as the set of points X_braid. The act of translation will be to copy and shift each of the sets by a fixed distance T. If the translated points, X_new = X_braid + T match the current braiding on the rope at that point X_new = X_current, then the translation, T, was symmetric. In our example this means that the translated "two cross and box" pattern matches the current pattern at that point on the rope. This translation can be applied as many times as we like, if our rope is long enough, and our new pattern will always match the braiding at that point. (See Figure 1.)

For rotational symmetry, using our flower pattern we can find the relation between the angle the flower is rotated and the number of petals on the flower. Start by marking one of the petals with a cross so the rotation can be seen. If there are n petals and each rotation takes us to the next petal, it will take n rotations for the all the petals to be marked, a 360-degree rotation. The angle of one rotation that moves the cross from one petal to the next is therefore 360/n.

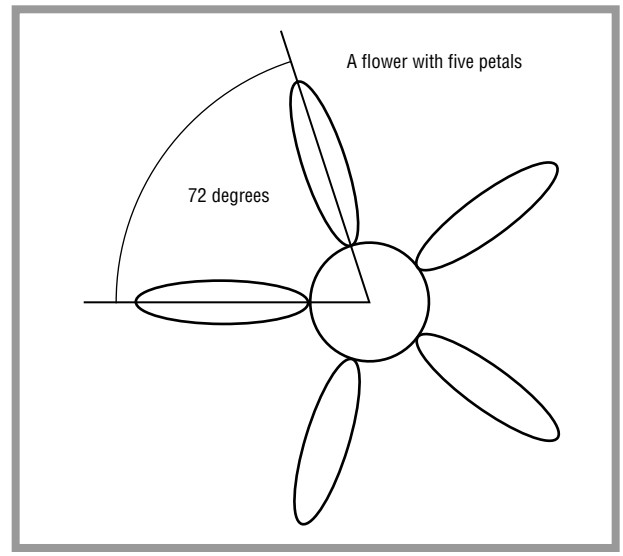


Figure 2.

As an example, think of a flower pattern with 5 evenly spaced petals. The smallest rotation that will leave the flower pattern unchanged is 360 / 5 petals = 72 degrees. So, if we wanted to draw a flower with five petals that has a rotational symmetry, each petal must be spaced exactly 72 degrees from the next. (See Figure 2.)

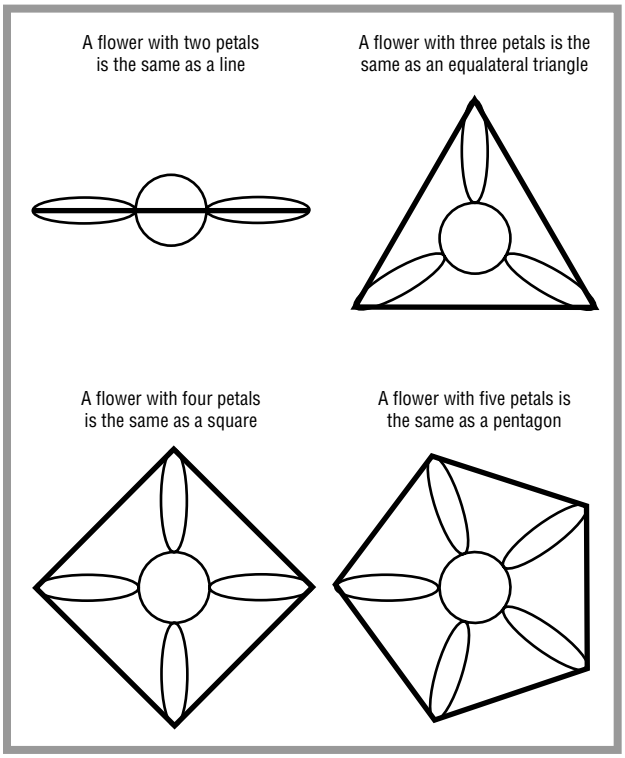


Figure 3.

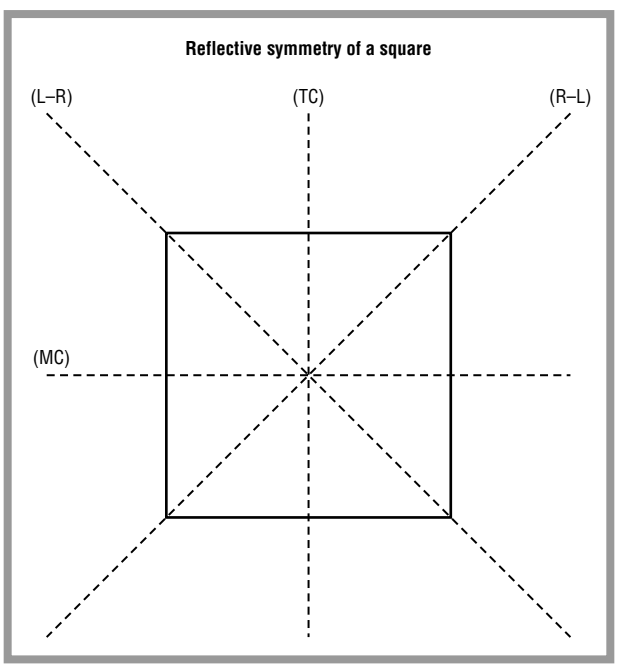


Figure 4.

Consider a flower pattern that is rotationally symmetric with four petals; this means that each petal will be spaced at 90 degrees from the next for the formula. Another shape that has four points that are each separated by 90 degrees around the center is a square. All of our flower patterns with rotational symmetry can be represented by geometrical shapes such as this. For example, a flower with two petals is identical to a line, with three it is identical to an equilateral triangle (a triangle where each side has an equal length), with four a square, with five a pentagon and so on. (See Figure 3.)

The operation for a reflection is defined by drawing a line that acts as a mirror. This is easier to see if it is done in stages for every point in our object. The first stage is to draw a line from a point through the mirror line; this is called the line of reflection and must cross the mirror line at exactly 90 degrees. The next step is to measure the distance along the line of reflection from the point to the mirror line. The reflection of the point is made by drawing a point on the opposite side of the mirror line at an equal distance along the line of reflection.

If an object is placed in front of the mirror line and we generate a number of reflected points behind the mirror line and join them up we simply have made a reflection of the object but this is not a symmetry of the object as the reflected points are not matched up with the original object. However, if we can place the mirror line in the center of the object and all the points match up we have a reflective symmetry, for example a mirror line drawn down the center of a photograph of a face almost shows this symmetry.

An example of perfect reflection symmetry is a square. Using the reflection operation, there are four lines that can be drawn that will keep the shape of a square. The first is from the top left hand corner to the bottom right hand corner (L-R), from the center of the top edge to the center of the bottom edge (TC), the top right corner to the bottom left corner (R-L) and along the center of the middle left to right edge (MC). Now, mark the top left corner with a cross as was done with the flower pattern to see the effect of each reflection. Under the (L-R) reflection the cross will not move, under (TC) the cross is reflected in the top right corner, under (R-L) the cross is reflected to the bottom right corner and under (MC) it is reflected to the bottom left corner. This is identical to the effect of the cross under rotation; our square has to be rotated four times to bring the cross back to the starting position and this is also the number of lines of reflection symmetry. (See Figure 4.)

Geometric objects with rotation symmetry, lines, triangles, and squares etc. have an identical symmetry to reflection. The number of reflection planes for a geometric shape is given by the number of rotations needed to make the object turn one full circle, 360 degrees. This is simply equal to the number of corners, or petals if we are using a flower shape. The angle from the center between two opposing corners is given by the rotation formula, 360/n.

An equilateral triangle, or a flower with three petals, has 3 lines of reflection and rotational symmetry every 45 degrees from the formula. A square, or a flower with four petals, will have 4 lines of reflection and rotational symmetry every 90 degrees and so on.

Real-life Applications

ARCHITECTURE

Nature is not the only one to use this tool, mankind uses it extensively as well. Look at most architecture and you will see symmetries used in the construction. The architect uses symmetry to distribute the forces in the building in a manageable way and for artistic reasons to give the building an appealing elegance. Many psychological studies have shown the human concept of symmetry is closely related to perceptions of beauty when humans evaluate shape as either beautiful or ugly.

The sun sets under the Arc de Triomphe in Paris, providing a display of symmetry for tourists and drivers on the Champs-Elysees. AP/WIDE WORLD PHOTOS. REPRODUCED BY PERMISSION.

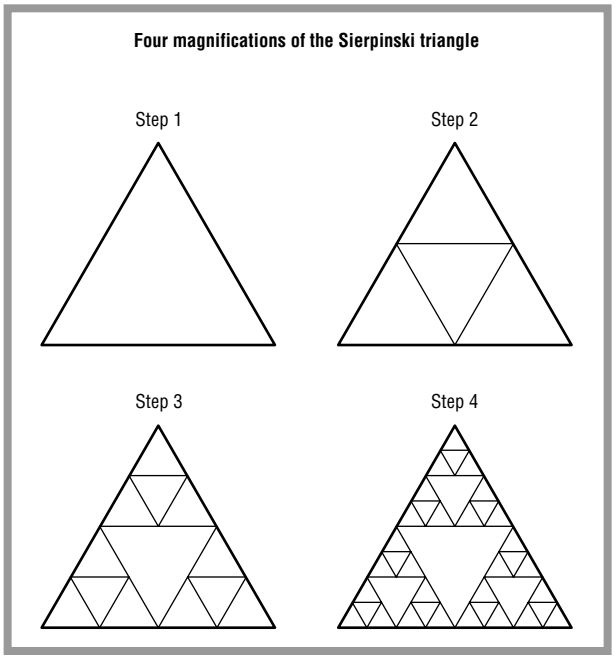


Figure 5: Sierpinski triangles.

Symmetry and Perceptions of Beauty

Symmetry has artistic importance as it is deeply embedded in our experience of the world and how it should look. It gives us strong feelings of how beautiful or ugly something is. Non-symmetric and symmetric shapes can affect us quite deeply, and these effects are exploited by modern artists. For an example Pablo Picasso often challenged perceptions of symmetry with a stunning effect.

FRACTAL SYMMETRIES

There is another special form of symmetry that is common in the natural world that is called fractal symmetry, or scaling symmetry. If a fractal is scaled up or down by a certain magnification it will look exactly the same as before. In nature this type of symmetry can be seen in many plants. A tree, for example, has a thick truck that divides into a number of branches; each branch then divides into a smaller branches and so on. It is possible to imagine that if one of these branches were scaled up it would look like a tree itself. This scaling symmetry is finding uses in many areas of science, such as weather patterns, the stock market and earthquake prediction, that are too complex to predict with normal means.

An example of a simple fractal shape is the Sierpinski triangle. To draw a Sierpinski triangle, start by drawing an equilateral triangle. Along the middle of each edge make a point. Each of these points is then joined by a line.

This divides the triangle into three smaller triangles. Repeat the operation along the middle edges of these triangles, only for the triangles that point upwards. This continues forever. (See Figure 5.) No matter how closely we magnify the triangle, we will always find that it is made from equilateral triangles exactly like the one we started with.

IMPERFECT SYMMETRIES

The symmetries seen in nature are rarely perfect. Consider the rotational symmetry of a flower; on closer inspection each petal will have a unique pattern and marks that define it from the rest. The translation symmetry of the marks on animals and the reflection symmetry of the human body is never quite perfect. A human face made with a perfect reflection symmetry by a computer will look odd. These imperfections probably arise due to perturbations in the natural replication process. It seems that these imperfections are important for giving natural objects unique identities.

SYMMETRIES IN NATURE

There are other forms of symmetry that can help us understand the way our world works. These symmetries are often highly abstract and mathematical but offer deep insight into science and nature. Our day-to-day experience of electricity and magnetism show that they exhibit very different characteristics, but under certain operations they can be shown to be the same, hence they are symmetric. This unified force is called electromagnetism. The study of this abstract symmetry has, amongst other things, allowed the development of radio technology used to carry mobile phone calls and television signals. The study of symmetries and their properties in mathematics is called group theory.

Key Terms

Reflection: The operation of moving all the points to an equal distance, on the opposite side of a line of reflection.

Rotation: The operation of moving all the points of an object through a fixed angle around a fixed point.

Translation: The operation of moving each point a fixed distance in the same direction.

Where to Learn More

Web sites

Picasso, Pablo. "Girl in Boat," 1938 <www.abcgallery.com/P/picasso/picasso205.html> (January 30, 2005).

Youngzones.org "Symmetry." <http://math.youngzones.org/symmetry.html> (January 21, 2005).

Tables

Overview

A table consists of a series of rows and columns that are used to organize various types of data.

Tables serve many uses in daily life, providing a teaching tool for basic math problems or easy access to the solutions to more complex equations, and displaying information in a logical format that enables anyone access to data they might otherwise be unable to determine themselves.

Fundamental Mathematical Concepts and Terms

Table headings, both along the top of columns and the start of rows, set the parameters for what sort of information the table will provide, and answers to each problem or question can be found by tracing each column and row to their point of intersection. The actual mathematical equation or work that has been done to provide each answer is done behind the scenes, and the table itself only displays the starting points and the solutions. In the case of highly complicated mathematics, the table provides the solutions to readers unable to perform the steps to find the answers on their own. Even if the equations are less complex, a table can save the time necessary to work out the solution by displaying the work that has been done earlier.

The amount of information provided by the table depends on the subject matter. Tables used for basic mathematical equations, such as addition or multiplication, generally cover basic numbers from 1 to 10, or 12, or else series of numbers, such as 10s or 100s. Complicated mathematical equations, such as logarithms, may be illustrated over many tables, each one covering a small number of problems. Other tables, where the mathematical application is less obvious, include the information relevant for that specific function. For instance, time tables for train departures would be based on the length of the day and the frequency of train trips.

A Brief History of Discovery and Development

The earliest records of the use of tables date back to Mesopotamia in approximately 3000 B.C., and the Sumerians, who used symbols to denote different goods and kept track of quantities of each item in table form on tablets. Between 2000 and 1600 B.C., the Old Babylonian period,

Types of tables

Tables fall into various categories, including mathematical, scientific, and astronomical. They can be used to communicate schedules, to translate weights and measures from one standard to another, or to convert currencies between countries. When designing a Web site, tables can help provide underlying structure as part of the coding of a Web page. Tables are available that convey various statistical information, including salary ranges for different jobs or for different regions; religious or political demographics by town, state, or country; types of disabilities and percentage of the population affected; health trends and life expectancies. Financial information is frequently presented in table format, from the transactions in your checking account, to interest rates, to income tax brackets. Tables provide ready-made data that enables you to get the information you need without extensive calculations or research.

Addition Table

+	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
0	0	1	2	3	4	5	6	7	8	9	10
1	1	2	3	4	5	6	7	8	9	10	11
2	2	3	4	5	6	7	8	9	10	11	12
3	3	4	5	6	7	8	9	10	11	12	13
4	4	5	6	7	8	9	10	11	12	13	14
5	5	6	7	8	9	10	11	12	13	14	15
6	6	7	8	9	10	11	12	13	14	15	16
7	7	8	9	10	11	12	13	14	15	16	17
8	8	9	10	11	12	13	14	15	16	17	18
9	9	10	11	12	13	14	15	16	17	18	19
10	10	11	12	13	14	15	16	17	18	19	20

Figure 1: Addition table.

two types of multiplication tables came into use. The first kind simply listed the multiples for a single number, while the second combined a series of these smaller tables on one tablet, illustrating a number of multiplication equations.

Single tables listed a principal number, denoted as p, then indicated the solutions for multiplying p by numbers 1 through 20. Mesopotamians operated on a base 60 system of numbers, and therefore could have calculated through 59p, but instead their tables went in increments of 10 following 20p, and to determine a number in between, one simply combined equations, so that 26p was the result of adding 20p with 6p. Combined tables included a variety of single tables on one tablet, with the individual tables generally written in descending order according to the primary number illustrated.

Examples of later tables were discovered in early Egyptian texts, listing the values of letters used as place holders in equations that contained fractions. Early Arabic astronomers utilized tables to keep track of variables used in calculating planetary positions necessary for astrology. Then in the eleventh century, Hebrew astronomers corrected errors in observations based on inconsistent calculations, and began recording constants in tables to ensure that all future observations were based on the same set of calculations.

As mathematical equations grew more complex, tables were used to record trigonometric functions, such as values for sine and cosine, and eventually logarithms and differential equations. In 1627, Johannes Kepler printed the first modern astronomical tables, the Rudolphine Tables, which provided accurate, complex planetary calculations based on longitude and latitude of the stars, previously not available.

Real-life Applications

MATH SKILLS

Tables are commonly used to teach basic math skills, such as addition and multiplication—equations that are eventually committed to memory. For addition, a series of numbers run across the top of the columns and down the start of the rows as headings. Then each number across the top is added to each number down the side and the result entered at the intersection of the column and row. The table is used to drill students in the basic addition skills. (See Figure 1.)

In multiplication, the numbers along the top of the columns are instead multiplied by those down the side of the rows, with the solution again placed at the intersection. This format allows certain patterns to become clear, often providing tricks that help students memorize the answers to the equations. For instance, when multiplying

Multiplication Table

×	0	1	2	3	4	5	6	7	8	9	10	11	12
0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	0	1	2	3	4	5	6	7	8	9	10	11	12
2	0	2	4	6	8	10	12	14	16	18	20	22	24
3	0	3	6	9	12	15	18	21	24	27	30	33	36
4	0	4	8	12	16	20	24	28	32	36	40	44	48
5	0	5	10	15	20	25	30	35	40	45	50	55	60
6	0	6	12	18	24	30	36	42	48	54	60	66	72
7	0	7	14	21	28	35	42	49	56	63	70	77	84
8	0	8	16	24	32	40	48	56	64	72	80	88	96
9	0	9	18	27	36	45	54	63	72	81	90	99	108
10	0	10	20	30	40	50	60	70	80	90	100	110	120
11	0	11	22	33	44	55	66	77	88	99	110	121	132
12	0	12	24	36	48	60	72	84	96	108	120	132	144

Figure 2: Multiplication table.

numbers by a factor of 9, as the first digit of the solution increases by one, the second digit decreases by one, so that $4 \times 9 = 36$, $5 \times 9 = 45$, $6 \times 9 = 54$, and so on. (See Figure 2.)

Similar tables can be used for both subtraction and division.

For more complicated math equations, tables list solutions that might normally require complicated calculations or even the use of a computer, or values for variables used in complex equations. Examples include trigonometry, logarithms, and differential equations.

OTHER EDUCATIONAL TABLES

Another table studied in school is the periodic table. Used in chemistry, this table displays the various elements according to abbreviation, their placement providing information such as atomic weight. Similar substances are grouped together, such as gases, liquids, and those elements that are synthetically crafted.

Probability and statistics findings can also be displayed using a table. Available data provides the material for headings, while the resulting odds for each combination fill in the body of the table. In genetics, this sort of information can be applied to the Punnet square. This small table illustrates the likelihood that various genetic traits will be inherited by a child, based on what genetic material each parent might donate, and whether that material is dominant or recessive. An example of this using two peapod plants, one tall carrying a recessive short gene (Tt) and one short (tt), where tall is the dominant trait, results in half of the second generation plants being tall and half of them short. (See Figure 4.)

CONVERTING MEASUREMENTS

Units of measurement vary from country to country, with some nations using metric measurements and others utilizing a standard of feet, yards, miles and so on. There are calculations that allow one to translate inches into centimeters or yards into meters, but tables that illustrate these transitions eliminate the need to recall the relationship between the two forms and provide a shortcut to performing the math. (See Figure 5.)

Tables can be used to illustrate not only equivalent measures of distance, but weight, liquid and solid capacity, or temperature, by converting pounds to kilos, quarts to liters, and degrees in Fahrenheit to degrees in Celsius. Fractions can be converted into decimals, with the table also illustrating the equivalent percentages. Cooking measurements, such as cups or tablespoons, may be listed as weights, in ounces, pounds, grams, and kilos, enabling a cook to translate recipes printed using American measurements into their own more familiar European measurements, or vice versa.

I	II	IIIb	IVb	Vb	VIb	VIIb	VIIIb			Ib	IIb	III	IV	V	VI	VII	0
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
H																	He
Li	Be											B	C	N	O	F	Ne
Na	Mg											Al	Si	P	S	Cl	Ar
K	Ca	Sc	Ti	V	Cr	Mn	Fe	Co	Ni	Cu	Zn	Ga	Ge	As	Se	Br	Kr
Rb	Sr	Y	Zr	Nb	Mo	Tc	Ru	Rh	Pd	Ag	Cd	In	Sn	Sb	Te	I	Xe
Cs	Ba	La*	Hf	Ta	W	Re	Os	Ir	Pt	Au	Hg	Tl	Pb	Bi	Po	At	Rn
Fr	Ra	Ac**	Rf	Db	Sg	Bh	Hs	Mt	Uun	Uuu	Uub		Uuq		Uuh		Uuo
Lanthanides *			Ce	Pr	Nd	Pm	Sm	Eu	Gd	Tb	Dy	Ho	Er	Tm	Yb	Lu	
Actinides **			Th	Pa	U	Np	Pu	Am	Cm	Bk	Cf	Es	Fm	Md	No	Lr	

Figure 3: The Periodic Table of the Elements.

Parent Pea Plants ("P" Generation)		Offspring ("F1" Generation)	
Genotypes: Tt x tt	Phenotypes: tall x short	Genotypes: 50% (2/4) Tt 50% (2/4) tt	Phenotypes: 50% tall 50% short

Figure 4: Punnet square. This small table illustrates the likelihood that various genetic traits will be inherited by a child, based on what genetic material each parent might donate, and whether that material is dominant or recessive.

Linear measure

1 mil	0.001 inch	0.0254 millimeter
1 inch	1,000 mils	2.54 centimeters
12 inches	1 foot	0.3048 meter
3 feet	1 yard	0.9144 meters
5.5 yards or 16.5 feet	1 rod (or pole or perch)	5.029 meters
1 mile	5,280 feet	1.6094 kilometers
40 rods	1 furlong	201.168 meters
8 furlongs	1 mile	1.6094 kilometers
3 miles	1 league	4.83 kilometers
	1 millimeter	0.03937 inches
10 milimeters	1 centimeter	0.3937 inch
10 centimeters	1 decimeter	3.937 inches
10 decimeters	1 meter	39.37 inches or 3.2808 feet
10 meters	1 decameter	393.7 inches or 32.8083 feet
10 decameters	1 hectometer	323.083 feet
10 hectometers	1 kilometer	0.621 mile or 3,280 feet
10 kilometers	1 myriameter	6.21 miles

Figure 5.

FINANCE

The financial industry makes use of tables as a way of conveying information for nearly every type of transaction. Checkbook registers are structured as very basic tables, with each row making up a separate transaction in your checking account, and the columns indicating what type of transaction has taken place—the writing of a check, a deposit of funds, the addition of interest, amounts, and whether the transaction was of a tax-deductible nature, with the final column keeping a running tally of the account balance. Statements for bank accounts are also structured as tables, with dates and transactions indicated in each row, and the columns separating deposits from withdrawals, interest, and an updated balance.

Banks list interest rates in tables, both for certificates of deposit (CDs) and other investment accounts, and for their advertised loan rates for cars or mortgages. In the

Local and Universal Time

Due to the shape and movement of Earth, the planet is divided into time zones based upon the amount of time it takes to make one complete rotation of the sun—approximately twenty-four hours. What time zone you are in determines the time of day at any given point, with your own time zone considered "local time." Tables list the various different time zones and enable you to determine what the corresponding time is in another part of world at a glance. An international timetable makes it possible to see the time of day anywhere in the world, simply by comparing the time zones. In addition, a detailed timetable will include information that accounts for the use of daylight savings time in the summer, as some parts of the world do not observe this manual change of the clocks, and those who do sometimes begin and end this period on different dates.

Coordinated Universal Time (UTC), formerly known as Greenwich Mean Time, sets the time standard for the entire world. It is essentially solar mean time, and moves at a variable rate to account for the fact that the planet does not rotate around the sun in precisely twenty-four hours, but rather is off by a few seconds that eventually add up. Because of this, UTC will occasionally leap ahead by several seconds in order to even out the time with the actual rotation of the planet. All other clocks in all other time zones are set to correspond to UTC. (See Figure A.)

Standard time zone conversions

Conversions from UTC to some US time zones: * = previous day

UTC (GMT)	Pacific standard	Mountain standard	Central standard	Eastern standard
00	4 pm *	5 pm *	6 pm *	7 pm *
01	5 pm *	6 pm *	7 pm *	8 pm *
02	6 pm *	7 pm *	8 pm *	9 pm *
03	7 pm *	8 pm *	9 pm *	10 pm *
04	8 pm *	9 pm *	10 pm *	11 pm *
05	9 pm *	10 pm *	11 pm *	12 mid
06	10 pm *	11 pm *	12 mid	1 am
07	11 pm *	12 mid	1 am	2 am
08	12 mid	1 am	2 am	3 am
09	1 am	2 am	3 am	4 am
10	2 am	3 am	4 am	5 am
11	3 am	4 am	5 am	6 am

Figure A.

case of CDs, the table will list a row for each locked-in time period, as interest rates differ based on the length of time the account is held. Corresponding interest rates are listed in the next column, followed by a column for APY, or the actual period yield that reports the amount of interest you would earn when accounting for compounding. For instance, a CD that is deposited for a preset period of six months might earn a 2.52% interest rate, which would result in an actual earned rate of 2.55%. In the case of money market accounts, while there is no set time period, some banks offer greater interest rates for larger deposits to encourage customers to keep more money in the institution. They display these advantages in table form, listing each deposit increment, followed by the interest rate given on that amount. (See Figure 6.)

Loan information is also listed in tables. Car loans vary both in time period and interest rate amount, so a table might list the average interest rate offered for a loan that is spread over 36 months, 48 months, and 60 months. Mortgage rates offer even more choices, including the length of the loan—anywhere from 15 to 30 years—and whether a loan rate is fixed or variable. Sometimes the

	Current	1 Month Prior	3 Month Prior	6 Month Prior	1 Year Prior
1-Month	1.25	1.25	1.13	1.06	0.89
6-Month	2.55	2.37	2.01	1.81	1.34
1-Year	3.28	3.01	2.63	2.33	1.71
2-Year	3.66	3.33	3.11	2.95	2.23
5-Year	4.25	4.09	3.97	4.03	3.49
3-Month Jumbo	2.15	2.06	1.82	1.63	1.19
6-Month Jumbo	2.82	2.62	2.26	1.98	1.46
1-Year Jumbo	3.49	3.18	2.81	2.5	1.85
2-Year Jumbo	3.9	3.52	3.27	3.12	2.34

Source: CDs provided by Bankrate.com

Figure 6: Interest and earning table.

Ideal Weight for Men

Height (in shoes)	Small Frame	Medium Frame	Large Frame
6'4"	162 to 176 lb	171 to 187 lb	181 to 207 lb
6'3"	158 to 172 lb	167 to 182 lb	176 to 202 lb
6'2"	155 to 168 lb	164 to 178 lb	172 to 197 lb
6'1"	152 to 164 lb	160 to 174 lb	168 to 192 lb
6'	149 to 160 lb	157 to 170 lb	164 to 188 lb
5'11"	146 to 157 lb	154 to 166 lb	161 to 184 lb
5'10"	144 to 154 lb	151 to 163 lb	158 to 180 lb
5'9"	142 to 151 lb	148 to 160 lb	155 to 176 lb
5'8"	140 to 148 lb	145 to 157 lb	152 to 172 lb
5'7"	138 to 145 lb	142 to 154 lb	149 to 168 lb
5'6"	136 to 142 lb	139 to 151 lb	146 to 164 lb
5'5"	134 to 140 lb	137 to 148 lb	144 to 160 lb
5'4"	132 to 138 lb	135 to 145 lb	142 to 156 lb
5'3"	130 to 136 lb	133 to 143 lb	140 to 153 lb
5'2"	128 to 134 lb	131 to 141 lb	138 to 150 lb

Figure 7: Generalized "healthy" weight ranges for men.

tables are used to illustrate trends in loan rates as well, listing a column for current rates and another for rates that were offered the previous week. If rates are rising, the table is an easy way to encourage customers to make a decision before prices go even higher by showing how much they have changed in a short period of time.

Other investments display necessary information in table form. The business section of most large newspapers includes the most recent closing prices for the various stock markets. These enormous tables go on for pages and list row after row of company names, followed by column after column of information regarding how each stock is trading, including the most recent price, the price the day before, percent that price has changed, high and low prices over the last year, and other pertinent information for investors. Anyone following the trends of a particular company has only to know what exchange it trades on and they can locate the stock information within the table.

The United States Federal Government provides their own set of financial tables to the public each year—the most recent tax tables. These tables are designed to help you determine how much income tax you owe on your previous year's salary. The tables are divided into sections based on your adjusted gross income (AGI), with each row indicating a salary range, such as between $18,000 and $18,050, followed by columns for the amount of tax owed based on whether you are filing as an individual, a married person filing with their spouse, a married person filing alone, or the head of your household. The tax tables provide tax payment information for salaries ranging from a single dollar to just under $100,000. For earned income over $100,000, there are additional tables that explain how to calculate taxes owed based on different, broader ranges of income.

The United States government also provides information about average government salaries, displaying their findings in tables. These figures are available for executive positions and mid-level jobs, or according to location, or for very specific posts, such as administrative law judges or law enforcement officials. Each row lists the pay grade category of the position, while the columns indicate the various raises available at that job level. General job statistics, not limited to government employees, are also available through census findings.

HEALTH

Certain basic health information often appears in tables. Healthy weight ranges for both men and women of varying heights are typically displayed in table format. Columns are labeled according to the size of the person's frame—small, medium, or large—and then each row lists a height. Weight ranges are listed for each body type. (See Figure 7 and Figure 8.)

Ideal Weight for Women

Height (in shoes)	Small Frame	Medium Frame	Large Frame
6'	138 to 151 lb	148 to 162 lb	158 to 179 lb
5'11"	135 to 148 lb	145 to 159 lb	155 to 176 lb
5'10"	132 to 145 lb	142 to 156 lb	152 to 173 lb
5'9"	129 to 142 lb	139 to 153 lb	149 to 170 lb
5'8"	126 to 139 lb	136 to 150 lb	146 to 167 lb
5'7"	123 to 136 lb	133 to 147 lb	143 to 163 lb
5'6"	120 to 133 lb	130 to 144 lb	140 to 159 lb
5'5"	117 to 130 lb	127 to 141 lb	137 to 155 lb
5'4"	114 to 127 lb	124 to 138 lb	134 to 151 lb
5'3"	111 to 124 lb	121 to 135 lb	131 to 147 lb
5'2"	108 to 121 lb	118 to 132 lb	128 to 143 lb
5'1"	106 to 118 lb	115 to 129 lb	125 to 140 lb
5'0"	104 to 115 lb	113 to 126 lb	122 to 137 lb
4'11"	103 to 113 lb	111 to 123 lb	120 to 134 lb
4'10"	102 to 111 lb	109 to 121 lb	118 to 131 lb

Figure 8: Generalized "healthy" weight ranges for women.

Expected life spans are also listed in tables. The numbers of years a person is expected to live is based on many factors, such as family history, eating and exercise regimens, whether or not they smoke, year they were born, and so on. However, by distilling all of this information, it is possible to come up with an average life expectancy for a person of a given age, and these are what are published in the tables. The information is useful to individuals planning for retirement, as it helps them determine how many years they will need to support themselves based on their investments and savings. Life insurance costs are also based on a person's age and how long they are expected to live.

TRAVEL

There are numerous uses for tables when traveling. Schedules are frequently illustrated in table format. In order to determine what train or bus to take, one must consult the timetable that lists the various train or bus stations and then shows the progression of each vehicle from its starting point to the destination. The schedules are determined based on distance between stations and the amount of time it takes to travel between points, but travelers can simply consult the schedule rather than working out the distances mathematically.

Travelers going to the beach may consult a table to determine when high or low tide will occur. The tables take into account time of day and day of the month, seasons, the cycle of the moon—everything that affects the transition of the tide. The equations are solved in advance and the results published in a table so that beach-goers have no need to understand the math necessary to determine the tide's movements.

Currency varies from country to country, and travelers need to determine how much things cost regardless of their location. Banks issue tables that estimate the conversion rate between currencies, providing a translation from one monetary denomination to another for basic round numbers, such as a hundred dollars and its equivalent in various other currencies. This enables travelers to determine quickly how much they are spending without performing complicated math equations.

Potential Applications

DAILY USE

Tables can be used to assist with many day-to-day tasks. Computer spreadsheet programs provide a template that can apply to various uses, whether or not they use mathematical equations.

If you are interested in keeping track of your finances—both how much money you earn from various

Single Life Expectancy
(For Use by Beneficiaries)

Age	Life Expectancy	Age	Life Expectancy
56	28.7	84	8.1
57	27.9	85	7.6
58	27.0	86	7.1
59	26.1	87	6.7
60	25.2	88	6.3
61	24.4	89	5.9
62	23.5	90	5.5
63	22.7	91	5.2
64	21.8	92	4.9
65	21.0	93	4.6
66	20.2	94	4.3
67	19.4	95	4.1
68	18.6	96	3.8
69	17.8	97	3.6
70	17.0	98	3.4
71	16.3	99	3.1
72	15.5	100	2.9
73	14.8	101	2.7
74	14.1	102	2.5
75	13.4	103	2.3
76	12.7	104	2.1
77	12.1	105	1.9
78	11.4	106	1.7
79	10.8	107	1.5
80	10.2	108	1.4
81	9.7	109	1.2
82	9.1	110	1.1
83	8.6	111 and over	1.0

Figure 9: Average life expectancies.

sources, and how much you spend on both necessities and pleasure—a simple table can provide the format. Each row can be numbered with a day of the month; the column headings might include such labels as job income, allowance, interest, and gifts for incoming funds, and car payment, gas, insurance, housing, food, utilities, clothing, and entertainment for expenditures. If you wish to become even more detailed, you can break out certain categories further. For example, entertainment might become movies, concert tickets, eating out, sports activities, and travel. At the end of the month, the spreadsheet program will enable you to calculate for different factors, such as money earned over the month, total money spent, money spent on necessities, and money spent on luxuries. This can be very useful if you are looking to cut back and save for a big-ticket item, such as a new stereo or a car. The table allows you to see at a glance where you have to spend your money, and where you might eliminate a few costs.

Tables do not always have to deal with numbers. Although they originally were used for mathematical purposes, the structure applies to many other things. Schedules are a prime example of this. While some schedules, such as those for transportation or attending classes, deal with times and dates, others might simply distribute tasks. A table could assign chores to different household members on a rotating basis, with each column labeled with a task and each row indicating the week. Each person determines when they are scheduled to do the dishes, vacuum, or take out the trash by finding their name at the intersection of the chore and the week it has been assigned to them.

Changeable table showing New York City's train schedule. REUTERS/CORBIS.

Where to Learn More

Books

Auth, Joanne Buhl. *Deskbook of Math Formulas and Tables: A Handy Reference to Math Formulas, Metric Tables, Terminology and Everyday Problem Solving.* New York, NY: Van Nostrand Reinhold Company, 1985.

Fitzpatrick, Gary L. *International Time Table.* Metuchen, NJ: The Scarecrow Press, Inc., 1990.

Grattan-Guinness, Ivor (editor). *The Norton History of the Mathematical Sciences: The Rainbow of Mathematics.* New York, NY: W.W. Norton and Co., 1998.

Web sites

Baby Steps through the Punnet Squares. <http://www .borg.com/~lubehawk/psquare.htm> (April 5, 2005).

Bloomberg.com. <http://www.bloomberg.com/markets/rates/> (April 3, 2005).

IRS Web site. <http://www.irs.gov/> (April 1, 2005).

Math.2.org. <http://www.math2.org.> (April 5, 2005).

Mesopotamian Mathematics. <http://it.stlawu.edu/~dmelvill/mesomath/> (April 2, 2005).

Periodic Table Styles. <http://chemlab.pc.maricopa.edu/periodic/styles.html> (April 5, 2005).

S.O.S. Mathematics Tables and Formulas. <http://www.sosmath.com/tables/tables/html> (April 5, 2005).

United States Office of Personnel Management. <http://www.opm.gov/oca/05tables/index.asp> (April 5, 2005).

Universal Time: UTC, UT1. <http://www.hartrao.ac.za/nccsdoc/slalib/sun67.hxt/node219.html> (April 5, 2005).

University of Michigan Health System. <http://www.med.umich.edu/1libr/primry/life15.htm> (April 5, 2005).

Key Terms

Array: A rectangular arrangement of numerical data in rows and columns, as in a matrix.

Logarithm: The power to which a base number, usually 10, has to be raised to in order to produce a specific number.

Trigonometry: A branch of applied mathematics concerned with the relationship between angles and their sides and the calculations based on them. First developed as a branch of geometry focusing on triangles during the third century B.C, trigonometry was used extensively for astronomical measurements. The major trigonometric functions, including sine, cosine, and tangent, were first defined as ratios of sides in a right triangle.

Topology

Overview

Topology is a branch of mathematics that studies the shapes of objects. More specifically, topology is concerned with how portions of an object do not change when a change in overall shape occurs. Two objects are considered to be the same if they can be changed to the other form without being cut or torn. An example is a bowl and a plate. At least in the imagination, it is possible to change a bowl into a flat plate by pressing down on the curved surface of the bowl (of course, you would not want to try this in your kitchen with your parents' best china).

Another way to visualize topology is to look at a map of the freeway system of a typical large city. Dozens of freeways intersect and fan out in different directions. Looking at the map in a topological manner would involve determining how the lines would connect to get a driver to a given destination. This has nothing to do with how far the journey would be, which would be more in the realm of geometry.

Looking at the freeway map from the viewpoint of topology, it would be much more appropriate to ask if the beginning and end of the journey were connected by the same road, or whether several roads had to be taken to reach the destination.

A topological freeway journey has to do with shape. That is the heart of topology; the shape of an object and how the object can be changed in shape without changing the *properties* of its shape. For example, if you hold a beach ball in your hands and pull with both hands to stretch the ball, the formerly spherical ball is now a football shape. This new shape is called an ellipsoid. Even though the sphere and the ellipsoid are different in shape, they have the same topology.

Other questions that can be asked about an object's topology include: are there holes in the object? Is the object hollow, and does the object have a limit, like a balloon, or does it reach to infinity.

Thinking about topology can be a bit confusing and mind-bending. Topology has been described as being "rubber-sheet geometry." That means that it is fine to shrink an object, or twist it or stretch it, because topology is not concerned with how close one molecule is to another. Changes like cutting, pasting, or puncturing, however, which take molecules from one part of the object and move them to another part of the object, are not part of studying topology.

An object's topology can be different depending on how it is viewed. Let us return to the dinner plate example to see an illustration of this concept. If a plate is lying on a dinner table and you look at it before you sit down,

A commuter passes a map of the London Underground (subway) system showing how the lines and stations connect (topological information), rather than distances, directions, or other geometric information. TOBY MELVILLE/REUTERS/CORBIS.

it appears like a circle. But, if you walk ten feet away and then look at the same plate, its shape may well be more like the elliptical football; the plate will look much wider than long. Two very different views are produced by an object whose shape has not changed.

Topological variations like the dinner plate example are one reason why learning how to draw can be a difficult process. Even though the shape of a plate from across the room is elliptical, the mind can still interpret a plate as being a circle. So, when drawing a table, the artist can actually fight against his or her brain telling him to draw a circle.

A dinner plate is a solid object, whose shape cannot be changed except by breaking it. But consider a piece of bread dough. The dough can be squeezed, rolled, flattened and stretched. A circle of dough can be changed to a rectangle, triangle, sphere, or other shapes. From a topological point of view, all the shapes of the piece of dough are the same. This is because to change from one form (a circle, for example) to another form (a triangle, for example) does not involve a rearrangement of the molecules that make up the object. A dough circle is the same as a shape like a dough triangle, as portions of the circle can be tugged outward to form the triangle. Likewise, a dough triangle is the same topology as a dough sphere, as the triangle can be rolled around the form a sphere.

Topology also involves the shape that can be created in space by the movement of an object. A good example of this is the hands of a clock. As the hour, minute, or second hands move from the 12 o'clock position around the dial and back to 12 o'clock, they create a circle.

Real-life Applications

VISUAL ANALYSIS

Anyone who has looked at a graph of scientific or other information has an appreciation for topology, even if they have not realized it. The relationships between two or more factors (for example, retail sales, day of the week, and age of shopper) can be detailed without the use of a graph. But, plotting the information in graph form, with the resulting hills and valleys, makes it much easier for people to interpret and to understand the information.

VISUAL REPRESENTATION

Topology is used to measure and evaluate magnetic and electrical fields. Topology can determine the shape of the fields. The changing shape of Earth's magnetic field varies the field's reactions to incoming solar particles and radiation. These changes often impact the performance of electronic navigation and communication devices.

Topology is also used in physics to describe string theory or to construct models of the shape of the universe that conform to observed data.

COMPUTER NETWORKING

Topology is critically important in designing computer network systems. The layout of workstations, hubs, switches, and servers constitutes the physical topology of the network and greatly impacts the capability and speed of transmitting data.

I.Q. TESTS

A number of psychology, medical, and I.Q. type tests utilize topological puzzles to assess visual and coordination skills as they relate to thinking or manipulative skills.

For example, the handcuff puzzle has been around for over 250 years. Two people, some lengths of rope, and some space to move around in are required. Each person uses a length of rope as handcuffs. The rope has loops at either end to fit over each hand. As well, each length of rope should be long enough to allow each person to move around without tripping.

As each person puts the rope handcuffs on, the lengths of rope are themselves looped together. The challenge is then to get themselves apart from each other without removing the handcuffs, or damaging the ropes.

But how can the ropes be separated from each other when they are looped together and when the ropes cannot be cut?

Here is the solution to the handcuff puzzle. Let us pretend that you are one of the handcuffed pair of people. You take the other person's rope and move it along yours until the rope is lying on one of your arms (make it the right arm). The other person's rope should not be wrapped around your rope. It should just be lying along your right arm. Now, take your left hand and reach through the handcuff around the right wrist. Grab the other person's rope, which is still lying along your right arm. Pull the rope through the handcuff over the right hand. Now let the rope go back through the handcuff. You should now be separated from the other person.

Gymnast competes for airspace with a soccer ball-shaped balloon. Topology helps mathematicians to characterize diverse shapes. AP/WIDE WORLD PHOTOS. REPRODUCED BY PERMISSION.

The basis of the trick is topology. As a bowl can be made to form a plate without altering the surface, so can the arrangement of the ropes be altered without the need for cutting or other damage.

MÖBIUS STRIP

At first it seems absurd; a strip of paper that has only one side. But that is the magic of the Möbius strip. Once again, at the heart of this "magic" is topology.

A Möbius strip is easy to make. A strip of paper is closed into a loop. Before the ends of the looped strip are taped together, however, one end is given a twist to produce a looped piece of paper with a half-twist in it.

Now comes the fun part. When a pen or pencil mark is made down the middle of the strip all the way around, the mark will be on both sides of the paper. It is the twist that does it, as it makes the mark change from one side of

the paper strip to the other side. In other words, a Möbius strip only has one side.

Likewise, the strip only has one edge. And, when the strip is cut down the middle, the result is one long strip instead of two separate strips.

The Möbius strip is not just a novelty. This form of topology used to be part of car engines, in the form of the fan belt. By having a twist in the belt, any strain imposed on the belt would be directed evenly over the whole surface of the belt.

Where to Learn More

Books

Gamelin, T.W., and R.E. Greene. *Introduction to Topology.* Mineola: Dover Publications, 1999.

Periodicals

Collins, G.P. "The Shapes of Space," *Scientific American* (July, 2004) 291: 94–103.

Web sites

Weisstein, E.W. "Topology."*MathWorld*" <http://mathworld.wolfram.com/Topology.html> (September 5, 2004).

Trigonometry

Overview

Trignometry is the study of relationships among the sides and angles of triangles, and derives its name from the Greek word for triangle, *trignon*. Real life uses of trigonometry include navigation, land surveying, global positioning system (GPS) applications, robotics, and the design of structures such as buildings and bridges. The height of Mt. Everest, the world's tallest peak, was calculated using trigonometry long before it was scaled by mountaineers.

Plane trigonometry involves trigonometric relationships that occur on a flat plane such as a piece of paper, whereas spherical trigonometry involves trigonometric relationships that occur on spheres such as planets. If the area being studied is small compared to the size of the sphere, it is often possible to obtain acceptably accurate results by using plane trigonometry for spherical problems. The grid systems that land surveyors use when laying out construction sites or locating property boundaries, for example, are formulated by assuming that Earth's curved surface can be represented by a series of flat planes. Although it is relatively easy to perform trigonometric calculations using points that lie on any one of the flat planes, it is very difficult to perform calculations using points that lie on more than one of the planes.

Fundamental Mathematical Concepts and Terms

MEASURING ANGLES

Plane angles are measured using wedge-shaped increments representing a fraction of a circle. The most common unit of angular measurement is the degree, which is denoted by the symbol °. A circle consists of 360°. Thus, an angle of 1° is 1/360 of a circle. For very accurate measurements of angles, degrees can be expressed in decimal form (for example, 10.5°) or in terms of minutes and seconds of arc. Each degree is divided into 60 minutes and each minute is divided into 60 seconds. In most branches of mathematics, including geometry and trigonometry, positive angles are measured counter-clockwise from the positive x-axis. Angles measured clockwise from the x-axis are negative.

The use of 60 rather than some other number, for example 10 or 100, to subdivide angles into minutes and seconds dates back to the Babylonian civilization, which arose around 1800 B.C. The Babylonians used a base 60 number system for their financial and scientific calculations, so it was natural for them to divide each degree into 60 seconds and each second into 60 degrees. The reason

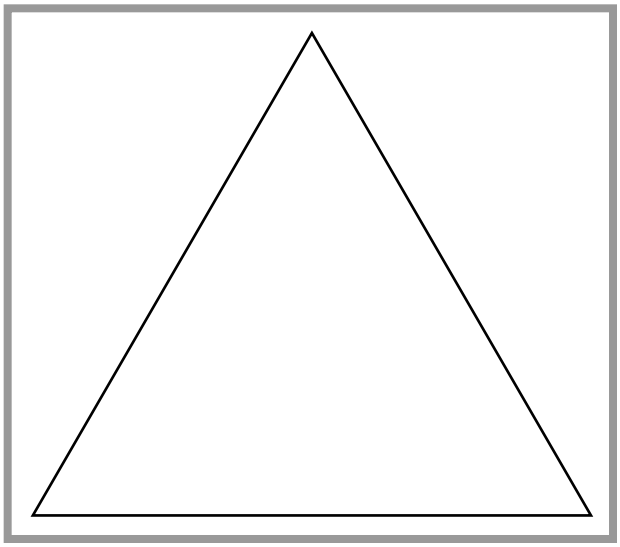


Figure 1: Equilateral triangle.

behind the division of circles into 360°, however, is less obvious. One explanation is that the Babylonians recognized that the sun follows a nearly circular path through the sky each year and that each Babylonian year consisted of 360 days. Thus, each degree in a geometric circle corresponded to one day in the Babylonian calendar. Another possible reason has been inferred from a Babylonian clay tablet unearthed in 1936, which describes geometric relationships within a circle circumscribed around a hexagon. The Babylonians knew that the perimeter of a hexagon is exactly six times the radius of a circumscribed circle, which led them to divide each of the six sides of the hexagon into 60 units, giving a total of 360° in a circle. The Babylonians also used relationship between hexagons and circumscribed circles to make a remarkably accurate estimate of the value of π. Regardless of the reason why circles were divided into 360°, it is a convenient integer because is divisible by 2, 3, 4, 5, 6, 8, 9, 10, 12, 15, 18, 20, 24, and 30. As such, fractions such as 1/2 or 1/5 of a circle can be represented by an integer number of degrees.

Degrees are not the only units that can be used to measure angles. In calculus and computer programs, angles are often measured using units called radians. There are 2π (or approximately 6.28) radians in a circle because the circumference of a circle with a radius of 1 is 2π. Thus, 1 radian represents one increment of length along the circumference of the circle, and is equal to approximately 57.3°. The need to quickly perform trigonometric calculations in battle led the German army to use the mil, which is short for milliradian, as a unit of angular measurement during World War II. It was subsequently adopted by other armies and is now the standard for angular measurement in military applications. A mil is 1/1,000 radian and is equal to the angle formed by a triangle 1,000 meters long and 1 m wide. This relationship applies only to small angles, so 1 radian cannot be defined in terms of a triangle that is 1 m long and 1 m wide. There are 6,283 or 2,000 π mils in a circle, so 1° = 17 mils. The German army and modern NATO armies use an approximate value of 6,400 mils in a circle, while the Soviet Union adopted an approximation of 6,000 mils in a circle. Binoculars, gunsights, and other instruments were calibrated and marked with angular measurements in mils in order to estimate distances. A truck that is 10 m long occupies 10 mils in the field of vision, and is approximately $10/10 \times 1{,}000 = 1{,}000$ m away. If the same truck occupies 20 mils, it is $10/20 \times 1{,}000 = 500$ m away.

A third form of angular measurement is the grad, which is a metric unit of angular measurement rarely used in the United States. One grad is defined as 1/400 of a circle, so it is slightly smaller than a degree. Instead of being divided into minutes and seconds, grads are divided into centigrads and milligrads.

Angles less than 90° are referred to as acute angles and those greater than 90° are referred to as obtuse angles. Angles that are exactly equal to 90° are known as right angles. The right in right angle is an outgrowth of the Latin word rectus, an adjective meaning correct or proper, that has found its way into English words such as direct, correct, erect, and rectify. A likely explanation is that a 90° angle is called a right angle because it is upright or erect, as in a wall that forms a right angle with the floor and ceiling in a house. This explains why there are no left angles. Right angles are sometimes described as orthogonal, which is derived from the Greek words for right (ortho) and angle (gonia).

TYPES OF TRIANGLES

Plane triangles can be classified according to the relative lengths of their three sides or the angles between the sides. In either case, three kinds of triangles are recognized. If classified according to the lengths of their sides, triangles are equilateral, isosceles, or scalene. If classified according to their angles, triangles are acute, right, or obtuse. Regardless of the triangle type, the sum of the three angles in a plane triangle must always add up to 180° or π radians. If two triangles are identical, they are said to be congruent. If they are of the same shape but different sizes, so that their angles are identical but the lengths of their sides are different, they are said to be similar.

Equilateral triangles have three sides of equal length and, as a consequence, three angles of equal size. (See Figure 1.) Although the lengths can be of any size, the

three angles in an equilateral triangle are all 60°. Isosceles triangles have two sides of equal length and a third side that is shorter than the other two. Two of the angles in an isosceles triangle are equal to each other and the third, located opposite the shortest side, is always smaller than the other two angles. Scalene triangles have three unequal sides and three unequal angles, with the largest angle opposite the longest side and the smallest angle opposite the shortest side. (See Figure 2.)

Acute triangles are defined by three acute angles, with no restrictions regarding the lengths of the sides. Thus, equilateral and isosceles triangles are acute triangles. Obtuse triangles contain one obtuse angle, and scalene triangles are necessarily also obtuse triangles. Because a triangle must consist of three angles that sum to 180°, a triangle cannot contain more than one obtuse angle. Right triangles are defined by the presence of one 90°, or right, angle. (See Figure 3.) The side opposite the right angle is known as the hypotenuse and, because of the right angle, the sum of the two remaining angles must always be 90°. Just as an obtuse triangle cannot contain more than one obtuse angle, a right triangle cannot contain more than one right angle.

Spherical triangles are formed by the intersection of three curved lines, or arcs, on the surface of a sphere.

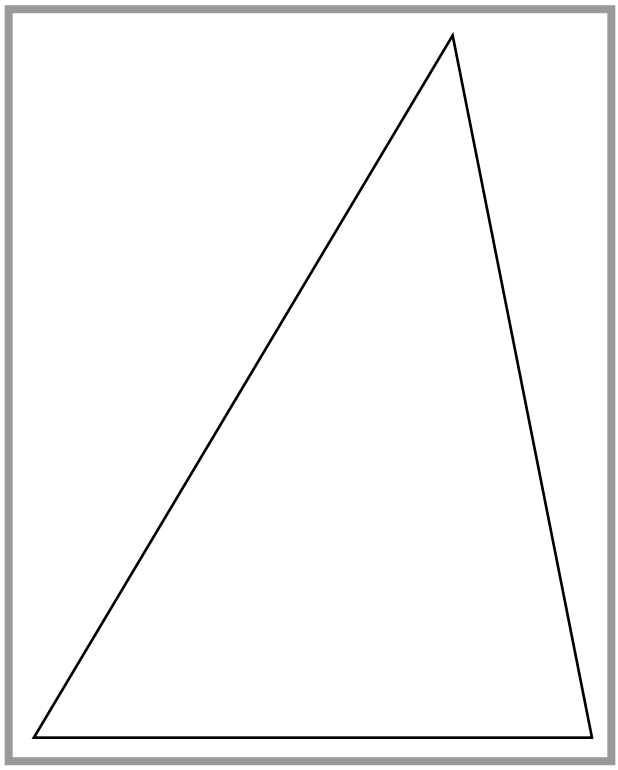


Figure 2: Scalene triangle.

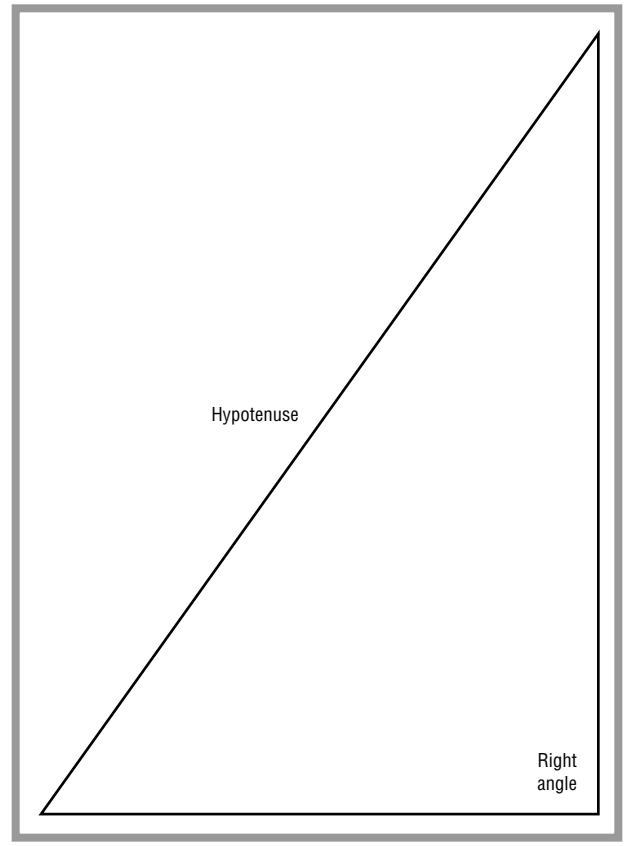


Figure 3: Right triangle.

Unlike the angles comprising a plane triangle, the angles inside of a spherical triangle do not always add up to a fixed value of 180°. Instead, they add up to a value between 180° and 540° (or π to 3π radians), and the difference between the sum of the angles and 180° is known as the spherical excess. As the surface area of the sphere becomes large relative to the size of the triangle, the spherical excess decreases towards zero and the spherical triangle becomes much like a plane triangle. This is why small areas of Earth's curved surface can be mapped as if they were planes. The second difference between spherical and plane triangles is that because a sphere wraps around on itself, the three intersecting arcs form one interior spherical triangle and one exterior spherical triangle. The sum of the angles of the outer spherical triangle always falls between 540° and 900° (3π and 5π radians). The third difference is that the lengths of the sides of spherical triangles can be measured in degrees as well as units of length.

PYTHAGOREAN THEOREM

Some of the most widely used real life applications of trigonometry are based on the Pythagorean theorem, which relates the lengths of the three sides of a right

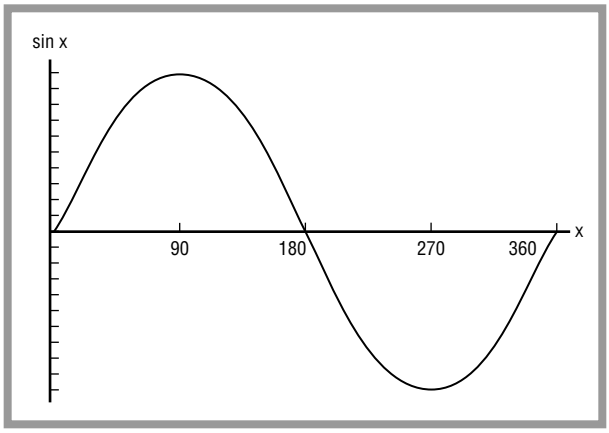


Figure 4: Sine curve.

triangle. Pythagoras (560–480 B.C.) was a Greek mathematician and philosopher who founded a school of religion and philosophy in the city of Croton. Pythagoras and his followers believed that geometric properties could always be expressed as ratios or products of whole numbers. They were troubled that the length of the hypotenuse of a right triangle is not a simple multiple of the lengths of the sides. Instead, the square of the hypotenuse is the sum of the squares of the two sides. If the lengths of any two sides of a right triangle are known, the Pythagorean theorem allows the length of the third side to be calculated.

If the length of the hypotenuse of a right triangle is C, and the lengths of the two sides are A and B, then the Pythagorean theorem is $A^2 + B^2 = C^2$. In the case of a triangle with sides of $A = 3$ and $B = 4$, $C = 5$ and the Pythagorean theorem is $3^2 + 4^2 = 5^2$. In this case the three lengths are related using only whole numbers. In the special case in which both $A = 1$ and $B = 1$, however, the theorem is $1 + 1 = 2$. Because $C^2 = 2$, $C = \sqrt{2}$, which is often referred to as Pythagoras's constant. The problem that faced Pythagoras and his followers was that $\sqrt{2}$ is an irrational number that cannot be expressed as a ratio of whole numbers such as ½ or ⅓, which violated their belief that geometric relationships must be expressible in terms of whole numbers. It is said that Hippasus, a follower of Pythagoras, was thrown overboard by other Pythagoreans when he proved during an ocean voyage that $\sqrt{2}$ is irrational. However disappointing it was to the Pythagoreans, the discovery of irrational numbers was an important step in the progress of mathematics that led to the acceptance of concepts such as π.

TRIGONOMETRIC FUNCTIONS

Real life applications of trigonometry almost always involve the use of three trigonometric functions that relate the sides and angles of right triangles. Those three functions are the sine, cosine, and tangent of an angle. To define the trigonometric functions, use the letter C to represent the length of the hypotenuse of a right triangle and the letters A and B to represent the lengths of the other two sides. It does not matter which of the two shorter sides is A and which is B, as long as C is the hypotenuse. Next, use the lower-case letters a, b, and c to represent the three angles. Assign the letters so that angle a is opposite side A, angle b is opposite side B, and angle c is opposite side C (which is the hypotenuse).

The sine of an angle, which is almost always abbreviated as sin, is the ratio of the side opposite the angle and the hypotenuse. In other words, $\sin a = A / C$ and $\sin b = B / C$. The actual numerical value of the sine function will depend on the size of the angle. For example the sine of 30°, which is abbreviated as sin 30°, is 1/2. One of the reasons that the sine function is an important tool for scientists, engineers, and mathematicians is that it is periodic, meaning that it repeats itself at regular intervals. It begins with $\sin 0° = 0$, increases until it reaches a peak at $\sin 90° = 1$, decreases to $\sin 180° = 0$ and then $\sin 270° = -1$, and then increases to $\sin 360° = 0$. This pattern repeats itself indefinitely every 360°. A plot of the sine function for angles ranging from 0° to 360° produces a wave-like line that is known as a sine curve. (See Figure 4.)

Another important trigonometric function is the cosine, which is defined as the ratio of the side adjacent to an angle and the hypotenuse. Using the same definitions as in the previous paragraph, the cosine of a, which is usually written as cos a, is B / C. Likewise, $\cos b = A / C$. The cosine function follows a curve that is identical in shape to a sine curve that has been shifted 90° to the left or right along the horizontal axis, starting with $\cos 0° = 1$ and decreasing to $\cos 90° = 0$ and $\cos 180° = -1$, then increasing smoothly to $\cos 270° = 0$ and $\cos 360° = 1$. Like the sine curve, the cosine curve repeats itself indefinitely.

The identical shapes of the sine and cosine curves can be expressed by the relationship $\sin \tau = \cos(\tau - 90°)$ or, equivalently, $\cos \tau = \sin(\tau + 90°)$, in which τ is any angle measured in degrees. The same relationship works for angles measured in radians if the 90° is changed to $\pi/2$ radians. Another way of writing the relationship between the sine and cosine of an angle is to use a variation of the Pythagorean theorem, which is that $\sin^2 \tau + \cos^2 \tau = 1$. Thus, if a person knows the sine of an angle then he or she can easily calculate the cosine, or vice versa.

Sine and cosine curves are important in many scientific and engineering problems involving waves in space or time. For example, a seismologist analyzing the vibrations

recorded during a large earthquake considers his or her seismogram to consist of many different sine and cosine waves added together to produce the complicated shaking. An oceanographer can analyze the waves traveling across a body of water using sine and cosine functions. Even Earth's topography can be simulated as a collection of sine and cosine waves added together. Although values of sines or cosines can be calculated by drawing a triangle to scale, measuring the lengths of the appropriate sides, and dividing, this is not an efficient approach for practical problems. An engineer who needs to draw a graph of a sine curve, for example, would have to draw many triangles in order to calculate the value of sin τ at enough values of τ to produce a smooth curve. Until computers and scientific calculators became widespread in the last half of the twentieth century, people who needed to find the values of trignometric functions looked them up in mathematical handbooks that contained tables of values for each function. Since then, however, so-called trig tables have virtually disappeared and values are almost always calculated using a calculator or a computer. Whereas most scientific calculators will allow their users to choose whether angles are specified in degrees, radians, or grads, computer languages generally require angular measurements to be specified in radians.

The third basic trignometric function is the tangent, which is the ratio of the sides opposite and adjacent to an angle. Again using the same variables as above, the tangent of a, or tan a, is A / B and tan b = B / A. As the size of the angle decreases, so does its tangent. The tangent of any angle that is an even multiple of 90° (for example, 0°, 180°, and 360°) is 0. As the size of the angle increases, the tangent increases until it reaches ∞ for angles that are odd multiples of 90° (for example, 90°, 270°, and 450°). The tangent of an angle can also be defined in terms of its sine and cosine, tan a = sin a / sin b. One of the reasons why angles measured in radians are useful in science and engineering is that if the angle is very small, the angle will very nearly equal to its tangent. For example, the tangent of an angle measuring 1/100 radian is 1/100. Because of this, some complicated equations involving the tangents of angles can be made much simpler. It is also the basis for the use of mils as angular measurements that can be used to easily calculate distances.

In addition to the three basic functions (sine, cosine, and tangent), there exist three reciprocal functions: cosecant, secant, and cotangent. The cosecant of angle a is csc a = 1/sin a, the secant of a is sec a = 1/cos a, and the cotangent of a is cot a = 1/tan a. Notice that the cosecant is the reciprocal of the sine and the secant is the reciprocal of the cosine, which can be confusing. The cosecant, secant, and cotangent can help to simplify complicated equations involving trigonometric terms. Two other trigonometric functions, the versed sine (versin a = 1 − cos a) and the exsecant (exsec a = sec a − 1), have fallen out of use.

Over the years many mnemonics have been proposed to help students remember which sides of a triangle are associated with each of the functions. To use the mnemonics, first take the first letter of each word in the following equations: Sine = Opposite / Hypotenuse, Cosine = Adjacent / Hypotenuse, and Tangent = Opposite/ Adjacent. This will form the combination SOHCAHTOA, which can be used to invent sentences consisting of words beginning with those letters, for example Some Old Horses Chase And Hunt 'Til Old Age or Silly Old Harry Caught a Herring Trawling Off America. Another approach is to use only the first letters of the names of the sides, which form the combination OHAHOA. Two common mnemonics involving those letters are Old Houses Always Have Old Attics and Oscar Has A Heap Of Acorns. Regardless of how clever they are, it may take more effort to memorize the mnemonics and their meanings than to simply memorize the basic definitions.

LAW OF SINES

The law of sines relates the sides and angles in any triangle, regardless of whether or not it is a right triangle. Again using upper case letters to represent the sides of a triangle and lower case letters to represent the angles opposite those sides, the law of sines is A / sin a = B / sin b = C / sin c = 2R, where R is the radius of a circle circumscribing the triangle. Any two of the terms separated by equal signs can be combined to perform trigonometric calculations. For example, if A = 1 cm and a = 30°, then 1 / sin 30° =2 R. Because sin 30° = 1/2, the law of sines requires that 1 / 1/2 = 2 R or, simplifying the equation, R = 1. Calculations can also be formed using only sides and angles. For example if A = 1 cm and a = 30°, as before, and b =50°, then the law of sines can be written as 1 cm / sin 30° = B / sin 50°. Knowing that sin 30° = 0.5000 and sin 50° = 0.7660, it follows that B = 1.53 cm.

A Brief History of Discovery and Development

Although the Babylonians developed the unit of angular measurement known today as a degree, knew many geometric techniques, and gave angular coordinates for stars, the Greek astronomer Hipparchus (180–125 B.C.) is generally known as the father of trigonometry. He constructed tables of chords, which are line segments joining two points along the circumference of a circle, and it is said that he wrote a 12–volume

treatise on chords. Although that work has never been found, it has been described by other Greek writers, and the series of volumes appears to have been the first written about trigonometry. The analysis of chords was an important development in the history of trigonometry because they are related to the sine and cosine functions. To illustrate this, consider a straight line connecting any two points along the circumference of a circle. Draw a line from one of the points to the center of the circle, and then from the center to the other end of the chord. This forms a triangle, and the length of the chord is related to the angle formed within the circle.

The Greek astronomer Menelaus (70–130 A.D.), about whom little is known, wrote a treatise on spherical trigonometry and its applications to astronomy. Like Hipparchus, he worked with chords rather than the modern trigonometric functions. He was a contemporary of Ptolemy (85–165 A.D.), another astronomer and author of a book most commonly known as the Almagest, which included tables of chords computed in %½° increments. Ptolemy also described how to use his tables of chords to solve trigonometric problems. Despite his great contributions to science and mathematics, Ptolemy is best remembered for his geocentric theory stating that the Sun and planets revolve around Earth. At about the same time, astronomers in India were using the precursor to the modern sine function rather than chords in their calculations. Unlike the modern sine function, the sine function developed in India was based on the length of one leg of a right triangle and not the ratio of the leg to the hypotenuse.

Muslim astronomers took the lead during the Middle Ages, building upon the work of their Greek and Indian predecessors. In particular, they began using all six modern day trigonometric functions (sine, cosine, tangent, secant, cosecant, and cotangent). Muslims continued to use lengths rather than ratios in their trigonometric functions, but also appear to have started using circles with a radius of 1, rather than the Babylonian value of 60, in their derivations. This produced the same values for the trigonometric functions that are used today. In addition to being used for astronomical calculations, their results helped the faithful to determine the direction to Mecca for prayers five times each day.

Latin translations of Arabic books did not make their way to Europe until the twelfth century. During the thirteenth century, German astronomer Georges Joachim proposed that trigonometric functions should be expressed as ratios rather than the lengths of lines. This was an important contribution because it meant that the values of the functions would depend only on the angle and not the actual lengths of the triangle legs. Like other scholars of his time, Joachim adopted the Latin name Rheticus. Subsequent work by European scholars included the development of many relationships involving multiple angles and powers of trignometric terms, which laid the groundwork for much of European science in the following centuries. Trigonometry also played an important role in Isaac Newton's invention of the calculus, which included ways to write trigonometric functions of an angle as infinite series involving powers of the angle. With the invention of the calculus, trigonometry was absorbed into the larger field of mathematics known as analysis.

Real-life applications

NAVIGATION

Pilots, mariners, and mountaineers all use trigonometric concepts to find their way from one point to another. In navigation, positive angles are measured clockwise from North and are known as azimuths. Azimuths convey direction, so they can range from 0° to 360°, and the word azimuth is sometimes used synonymously with the word heading. The azimuth of a line running from south to north is 0° and the azimuth of a line running from north to south is 180°. This distinction is critical in navigation. In other applications, it may not be critical to distinguish the direction. For example, it does not matter whether the boundary of a country runs from north to south or south to north.

A related term, which was in common use before computers and calculators existed, is bearing. Like an azimuth, a bearing conveys the direction of a line. The difference is that a bearing is always an acute angle measured from north or south. The reason for this was that the trigonometric tables necessary for navigation calculations contained values ranging from 0° to 90°, so it was impossible to look up a value for, say, an azimuth of 140°. Instead of using an azimuth for the obtuse angle of 140° measured from north, a navigator would use its acute supplementary angle, which is $180° - 140° = 40°$. In order to avoid confusing this value with an azimuth of 40°, the bearing includes the axis from which the acute angle is measured (south in this case) and the direction in which it was measured (towards east in this case). Thus, the complete expression for the bearing would be south 40° east, or S40°E. One way to ensure that an azimuth is never confused with a bearing is to always write an azimuth using three digits and a bearing with two. Thus, an azimuth of 40° can be written 040° to avoid confusion

with a bearing of S40°E or N40°E. Any angle greater than 90° cannot be a proper bearing.

In real-life problems involving travel over large distances, Earth's curvature becomes important and spherical, rather than planar, and trigonometry must be used. Coordinates for navigation over long distances are given in terms of latitude and longitude, which are angular measurements. Trigonometry is used to calculate the distance between the starting and ending points of a journey, taking into account that the path follows the surface of a sphere and not a straight line. The latitude and longitude of waypoints along a journey can also be calculated using trigonometry.

Navigation on Earth is complicated by the fact that the North Magnetic Pole, to which compass needles are attracted, does not coincide with the North Geographic Pole. The North Magnetic Pole is located in far northern Canada. For very approximate navigation, for example if a hiker wants to know if she is generally headed north or south, the fact that the geographic and magnetic poles are different does not make much difference. For any kind of precise navigation or mapmaking, however, the difference is important. The difference between true north, which is the direction to the North Geographic Pole, and magnetic north, which is the direction to the North Magnetic Pole, is known as magnetic declination. It is shown as an angle on topographic maps and navigational charts. Magnetic north is about 20° east of true north in the northwestern United States and about 20° west of true north in the northeastern United States. The line of zero declination runs through the Midwestern part of the country. In other areas of the world, the magnetic declination can be as great as 90° east or west in the far southern hemisphere. The North Magnetic Pole moves from year to year as a consequence of Earth's rotation, so the magnetic declination also changes over time. Government agencies responsible for providing navigation aids track the movement of the North Magnetic Pole, and maps are continually revised to reflect changing declination. Measurements by the Canadian government show that the North Magnetic Pole moved an average of 25 miles (40 km) per year between 2001 and 2005.

A simple trigonometric calculation illustrates the error that can occur if magnetic declination is not taken into account. The distance off course will be the distance traveled multiplied by the sine of the magnetic declination. In an area where the magnetic declination is 20°, therefore, a sailor following a course due north would find herself 34 kilometers (21 mi) off course at the end of a 100-kilometers (62 mi) trip. The longer the distance traveled, the farther off course the traveler will be. If the magnetic declination is only 10°, however, the error will be 100 km $\times$ sin 10° = 17 km (11 mi).

VECTORS, FORCES, AND VELOCITIES

Vectors are quantities that have both direction and magnitude, for example the velocity of an automobile, airplane, or ship. The direction is the azimuth in which the vehicle is traveling and the magnitude is its speed. Using trigonometry, vectors can also be broken down into perpendicular components that can be added or subtracted. Take the example of a ferry that carries cars and trucks across a large river. If there are ferry docks directly across from each other on opposite banks of the river, the captain must steer the ferry upstream into the current in order to arrive at the other dock. Otherwise, the river current would push the ferry downstream and it would miss the dock. If the velocities of the river current and the ferry are known, then the captain can calculate the direction in which he must steer to end up at the other dock. The velocity of the river current forms one leg of a right triangle and the velocity of the ferry forms the hypotenuse (because the captain must point the ferry diagonally across the river to account for the current). If the current is moving at 5 km/hr and the ferry can travel at 12 km/hr, the angle at which the ferry needs to travel is found by calculating its sine. In this case, the sine of the unknown angle is 5 / 12 = 0.4167. The angle can then be determined by looking in a table of trignometric functions to find the angle that most closely matches the calculated value of 0.4167, by using a calculator to calculate the sines of different angles and comparing the results, or by using the arc sine (asin) function. Each of the trigonometric functions has an inverse function that allows the angle to be calculated from the value of the function. In this case, the answer is asin 0.4167 = 25°. In other words, the captain must point his ferry 25° upstream in order to account for the current and arrive at the dock directly across the river.

Another real-life application of vectors and trigonometry involves weight and friction. Automobiles and trains rely on friction to move uphill or remain in place when parked, and friction is required in order to hold soil and rock in place on steep slopes. If there is not enough friction, cars will slide uncontrollably downhill and landslides will occur. Even if a car is traveling downhill, friction is required to steer. In the simplest case, the traction of a vehicle or the resistance of a soil layer to landsliding depends on three things: the weight of the object, the coefficient of friction, and the steepness of the slope. The weight of the object is self-explanatory. The coefficient of friction is an experimentally measured value that depends on the two surfaces in contact with each other and, in some cases, temperature or the rate of movement. The value used before movement begins, for example between the tires of a parked car and the pavement or a

soil layer that is in place, is known as the static coefficient of friction. Once the object begins moving, the coefficient of friction decreases and is known as the dynamic coefficient of friction. Some typical examples of coefficients of friction are 0.7 for tires on dry asphalt, 0.4 for tires on frosty roads, and about 0.2 for tires on ice. The coefficients of friction for soils involved in landslides can range from about 0.3 to 1.0, with most values around 0.6.

Because weight is a force that acts vertically downward, trigonometry must be used to calculate the components of weight that are acting parallel to the sloping surface. The frictional force resisting movement parallel to the slope is $\mu \times w \times \cos \tau$, where w is the weight, μ is the coefficient of friction, and τ is the slope angle. The component of the weight acting downslope is $w \times \sin \tau$. Division of the frictional resisting force by the gravitational driving force gives the expression $\mu / \tan \tau$. If the result is equal or greater than 1, the car or soil layer will not slide downhill. If it is less than 1, then downhill sliding is inevitable. If the coefficient of friction for tires on dry asphalt is 0.7, then parked cars will slide downhill if the slope is greater than 35°. If the road is covered with ice, however, the coefficient of friction is only 0.2 and cars will slide downhill on slopes greater than 11°.

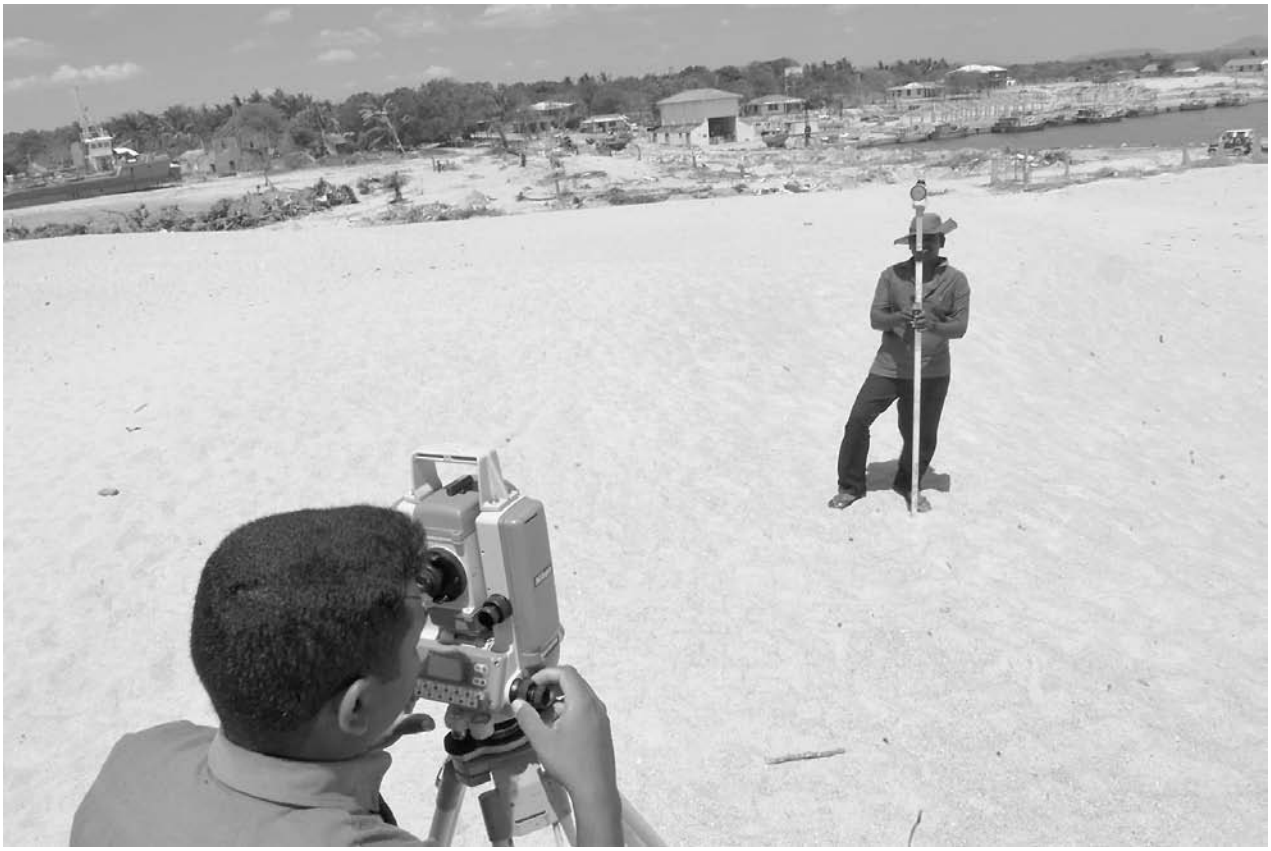
SURVEYING, GEODESY, AND MAPPING

Land surveyors and cartographers make extensive use of trigonometry in their work. Land surveyors are responsible for establishing official property boundaries and locations, for example the legal location of a piece of property being bought or sold. They also perform topographic surveys that depict the elevation of Earth's surface and important features such as stream channels, roads, utility lines, and buildings. Location maps showing the locations of oil, gas, and water wells also fall within the scope of land surveying. Because their work is used in legal transactions such as real estate sales and applications to drill wells, land surveyors are licensed by government agencies and must have a good knowledge of trigonometry. Geodetic scientists perform work that is similar to that done by land surveyors, but over much larger distances and in some cases with much greater demands for accuracy. They are responsible for establishing the networks of known points that land surveyors rely upon in their daily work. Cartographers use information provided by surveyors, geodetic scientists, and others to produce maps. One of the great challenges of cartography is the development of techniques to represent the three dimensional surfaces of Earth and other planets on two dimensional pieces of paper or computer monitors.

In order to determine the exact location of a property line or other feature, a surveyor begins at a point with a known location, known as a point of beginning. These are generally small metal discs or monuments established by government agencies such as the U.S. Geological Survey or the U.S. Coast and Geodetic Survey, and for which the location has been carefully determined in advance. The discs marking a point of known location are called benchmarks. Surveyors determine locations by using optical and electronic instruments to accurately and precisely measure angles and distances from a benchmark to the points for which locations must be determined. The angles and distances are plotted to create a map from which the locations of new points, for example the corners of a rectangular piece of property, can be calculated.

One of the techniques used by surveyors is triangulation, which allows them to determine the location of a point without actually occupying it. This is done by accurately measuring the length of a line between two known points, which is known as a baseline. The azimuth to the unknown point is measured from each of the known points, forming an imaginary triangle. The surveyor knows the length of one leg of the triangle and two of its angles, and can use trigonometry to calculate the lengths of the remaining two legs. This provides the location of the unknown point. The surveyors then move their instruments and use the newly calculated lengths as the sides of two more triangles, repeating the process to create a network of benchmarks. The locations of some or all of the points are calculated more than once, using different baselines each time, in order to improve the accuracy of the survey.

One of the greatest surveying projects of all time was the Great Trigonometric Survey, which was undertaken to map India when it was a British colony during the nineteenth century. Because of the great distances involved, the surveyors used specially manufactured theodolites, which are surveying instruments designed to measure angles. Surveyors look through a telescope at the center of the theodolite to align it with their target, and then read angles measured both horizontally and vertically. The theodolites manufactured for the Great Trigonometric Survey used circles 36 inches in diameter (the larger the circle, the more accurately angles can be measured) that were read through microscropes to achieve extremely high accuracy. Theodolites used for ordinary surveying, by way of comparison, have circles just a few inches in diameter. After the measurements were made, elaborate trigonometric calculations were made by hand in order to calculate the horizontal and vertical distances between points. In 1852, the Great Trigonometric Survey measured the height of a mountain known as Peak XV. Its



Surveyors use basic trigonometry while taking measurements at Kirinda, about 140 miles (225 km) southeast of Colombo, Sri Lanka. AP/WIDE WORLD PHOTOS. REPRODUCED BY PERMISSION.

height and location were calculated using a triangulation from six different locations, each of them at least 100 miles (160 km) from the peak. The results showed it to be the highest mountain on Earth, with a summit elevation of 8,850 m (29,035 ft). In 1856 Peak XV was renamed in honor of George Everest, the previous Surveyor General of India. It would be more than 100 years after its discovery before Sir Edmund Hillary and Tenzing Norgay reached the summit of Mt. Everest in 1953.

Even when global positioning system (GPS) receivers are used to determine the locations of unknown points, the locations of known points are used to increase accuracy. This is done by placing one GPS receiver over a known point such as a benchmark and using a second receiver at the point for which a location must be determined. In the United States, one of the known points might be a continuously operating reference station, or CORS, operated by the government and providing data to surveyors over the internet. Data from the two receivers are combined, either in real time or afterwards by post-processing, to obtain a more accurate solution that can be accurate to a millimeter or so. Although it may not be obvious because the calculations are performed by microprocessors within the GPS receivers and on computers, they require extensive use of trigonometric functions and principles.

Once the locations of points or features are determined, they must be plotted on a map in order to be visualized. If the area of concern is relatively small, the map can be constructed using an orthogonal grid system of perpendicular lines measuring the north-south and east-west distance from an arbitrary point. If the area to be mapped is large, however, then trigonometry must be used to project the nearly spherical surface of Earth onto a flat plane. Over the centuries, cartographers and mathematicians have developed many specialized projections involving trigonometric functions. Some are designed so that angles measured on the flat map are identical to those measured on a round globe; some are designed so that straight-line paths on the globe are preserved as straight lines on the planar map.

COMPUTER GRAPHICS

Both two- and three-dimensional computer graphics applications make heavy use of trigonometric relationships and formulae. Rotating an object in two dimensions, for example a spinning object in a video game or the text in an illustration, requires calculation of the sine and cosine of the angle through which the object is being rotated. Graphics objects are typically defined using points for which x and y coordinates are known. In some cases, the points may represent the ends of lines or the vertices of polygons. Many computer programs that allow users to rotate objects require the user to enter an angle of rotation or use a graphics tool that allows for freehand rotation in real time. Each time a new angle is entered or the mouse is moved to rotate an object on the screen, the new coordinates of each point must be quickly calculated behind the scenes.

Rotation of graphical objects in three dimensions is much more complicated than it is in two dimensions. This is because instead of one angle of rotation, three angles must be given. Although there are several different conventions that can be used to specify the three angles of rotation, the one that is most understandable to many people is based on roll, pitch, and yaw. These terms were originally used to describe the three kinds of rotation experienced in a ship as it moves across the sea, and were adopted to describe the motion of aircraft in the twentieth century. Roll refers to the side-to-side rotation of a ship or aircraft around horizontal axis. An aircraft is rolling if one of its wings goes up and the other goes down. Pitch refers to the upward or downward rotation of the bow of a ship or the nose of an aircraft. As the bow or nose goes up, the stern or tail goes down and vice versa. The final component of three-dimensional rotation is yaw, which refers to the side-to-side rotation of the nose or bow around a vertical axis. Just as in two-dimensional graphics, the simulation of three-dimensional rotation by a computer program requires that trigonometric functions be calculated for each of the three angles and applied to each point or polygon vertex. Three-dimensional graphics are also more complicated because the shape of each object being simulated must be projected onto a two dimensional computer monitor or other plane, just as Earth's spherical surface must be projected to make a map.

CHEMICAL ANALYSIS

Chemists rely on trigonometry to analyze unknown substances using methods such as Fourier transform spectroscopy. Spectroscopes are instruments that break down the electromagnetic radiation absorbed or emitted by a substance into a collection of component wavelengths known as a spectrum. Spectroscopes attached to telescopes, for example, allow astronomers to learn the chemical composition of distant stars by analyzing the color of their light. In the laboratory, chemists use a variety of spectroscopic techniques to determine the chemical composition of substances. Invisible forms of electromagnetic radiation such as infrared radiation can also be used for spectroscopy.

In conventional spectroscopy, the wavelength of electromagnetic radiation to which a sample is subjected is varied over a period of time. The result is a graph showing the response of the sample to different wavelengths of radiation. Fourier transform spectroscopy is different because the sample is subjected to many different wavelengths at once, which produces a complicated combination of responses to those many wavelengths of radiation. Its advantage is that Fourier transform spectroscopy is much quicker than conventional spectroscopy. Its disadvantage is that a series of complicated calculations known as a Fourier transform, which is based on sines and cosines, must be performed in order to make the results useful. Although the precursor of the Fourier spectroscope was invented in 1880, real life Fourier transform calculations are so time consuming that practical applications of Fourier transform spectroscopy were limited until digital computers became common in the 1950s. The subsequent discovery of an especially efficient computational method known as the fast Fourier transform, or FFT, in 1956 was a major advance that led to the practical development of Fourier transform spectroscopy. Fast Fourier transforms have since become an important computational tool for digital audio processing, digital image processing, and seismic data processing.

Potential applications

Trigonometric principles and calculations have applications in virtually every discipline of science and engineering, so their importance will continue to increase as technology continues to grow in importance. Global positioning system (GPS) receivers embedded in cellular telephones, vehicles, and emergency transmitters will allow lost travelers to be located and criminal suspects to be tracked. Fast Fourier transforms will help to advance any kind of computation involving waveforms, including voice recognition technologies. Trigonometric calculations related to navigation will also become even more important as global air travel increases and the responsibility for air traffic control is shifted from humans to computers and GPS technology.

Key Terms

Chord: A straight line connecting any two points on a curve.

Hypotenuse: The longest leg of a right triangle, located opposite the right angle.

Where to Learn More

Books

Downing, Douglas. *Trigonometry the Easy Way*, 3rd Ed. Hauppauge, New York: Barron's Educational Series, 2001.

Keay, John. *The Great Arc: The Dramatic Tale of How India was Mapped and Everest was Named*. New York: HarperCollins, 2000.

Maor, Eli. *Trigonometric Delights (Reprint edition)*. Princeton, New Jersey: Princeton University Press, 2002.

Schneider, Philip, and David H. Eberly. *Geometric Tools for Computer Graphics*. San Francisco: Morgan Kaufman, 2002.

Web sites

The Institute of Navigation. "Navigation Education Materials." August 21, 2000. <http://www.ion.org/satdiv/education.cfm> (May 9, 2005).

Stern, David P. "(M-7) Trigonometry—What is it good for?" November 25, 2001. <http://www-istp.gsfc.nasa.gov/stargaze/Strig1.htm> (May 5, 2005).

University of Cambridge. "Our Own Galaxy: The Milky Way." Cambridge Cosmology. 2005. <http://www.damtp.cam.ac.uk/user/gr/public/gal_milky.htm> (October 19, 2002).

Vectors

Overview

A vector is a mathematical object that contains two or more numbers in an ordered set. For example, [5 6 8] is a vector. Vectors containing two or three numbers (that is, vectors in two or three "dimensions") can be drawn as arrows. Vectors describe things that have more than one measurable feature: an arrow, for instance, has a certain length and points in a certain direction. Vectors are used in physics, medicine, engineering, and animation to describe forces, positions, speeds, changes in speed, electric and magnetic fields, gravity, and many other physical quantities. Vectors are also used in the field of linear algebra, along with the arrays of numbers called matrices, to stand for more abstract quantities. The rules for doing math with vectors are called "vector algebra."

Fundamental Mathematical Concepts and Terms

TWO-DIMENSIONAL VECTORS

A vector containing two numbers is termed a "two-dimensional" vector. A two-dimensional vector can be drawn as an arrow. To see how a pair of numbers can stand for an arrow (or the other way around), picture an arrow 25 centimeters (cm) long that has been drawn near one corner of a piece of paper. (See Figure 1.)

With the arrow drawn as in Figure 1, its base and tip are 3 cm apart as measured along the bottom edge of the

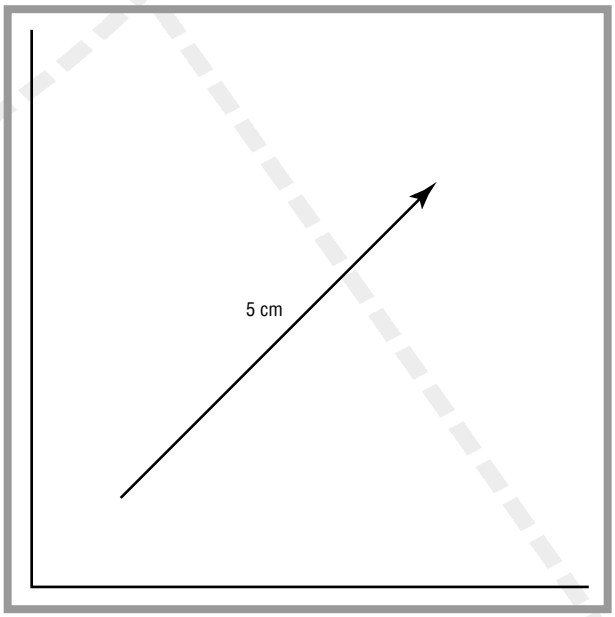


Figure 1. The arrow is not drawn to scale (that is, it is not exactly 5 cm long).

paper and 4 cm apart as measured along the side. (See Figure 2.)

The arrow can therefore be described by the vector [3 4]. Since this arrow is 5 cm long and is drawn at a 53° angle to the bottom edge of the paper, it could also be described by writing down its length and its angle: 5 cm, 53°. It cannot be described using fewer than two numbers. It is therefore called a "two-dimensional" vector.

Having more than one dimension—that is, consisting of more than one number—is what makes a vector different from a plain number (also called a "scalar"). A scalar can describe a measurement that consists of a single number, like the mass of a rock, but anything that pushes or points in a certain direction is best described by a vector. The force of a dancer's shoe against a floor, for example, is a vector because it has both a strength or "magnitude"—how hard the shoe is pushing against the boards—and a direction. This force can either be pictured as an arrow pushing against the sole of the shoe or written out in numbers. The push of a rocket motor, the speed and direction of a moving object, the strength and direction of an electric or magnetic field—all these things, and many more, are best treated as vectors.

THREE-DIMENSIONAL VECTORS

Many things in the real world cannot be described using two-dimensional vectors because they exist in three-dimensional space, namely, the ordinary space in which we live and move. Imagine, for example, an arrow placed inside a glass box. The arrow is too big to lie down in the box, but it can just fit with its base in one corner and its point in another. Drawing the edges of the box as dashed lines, we develop the image in Figure 3.

The length of this arrow and the direction it is pointing can be described by three numbers (also called "dimensions"). For example, if the glass box is 4 cm wide, 4 cm deep, and 8 cm tall, then writing down these numbers tells us exactly where the two ends of the arrow are. Writing down these three numbers gives us a numerical vector, [4 4 8]. In real life, many forces must be described using three-dimensional vectors like this one. For example, the force of a dancer's shoe against the floor is a three-dimensional vector because it might push on at the floor at any angle.

To write vectors in higher dimensions, all one has to do is write more numbers between the brackets, like [4 4 8 10 2]. Although it is not possible to draw pictures of vectors in higher dimensions, such vectors have many uses in physics, economics, and other fields.

THE MAGNITUDE OF A VECTOR

A vector's "magnitude" is basically its size. If a vector connects two points on a flat surface or in a three-dimensional space, then its magnitude gives the distance between those two points—the number we would measure if we stretched a measuring tape between them. If a vector

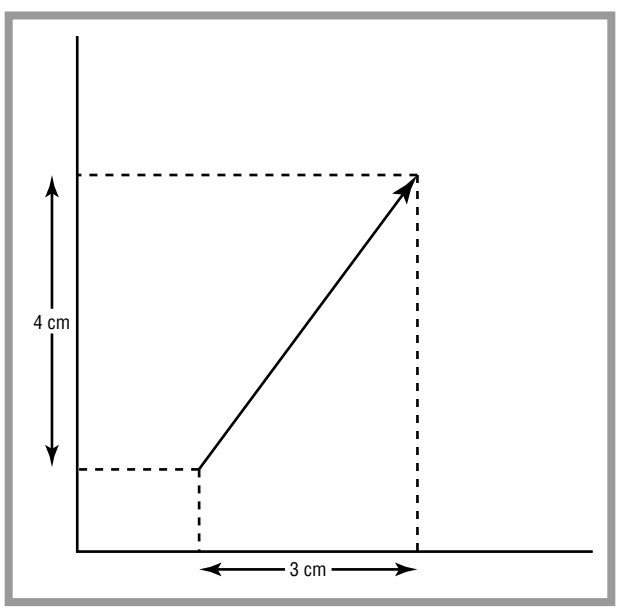


Figure 2.

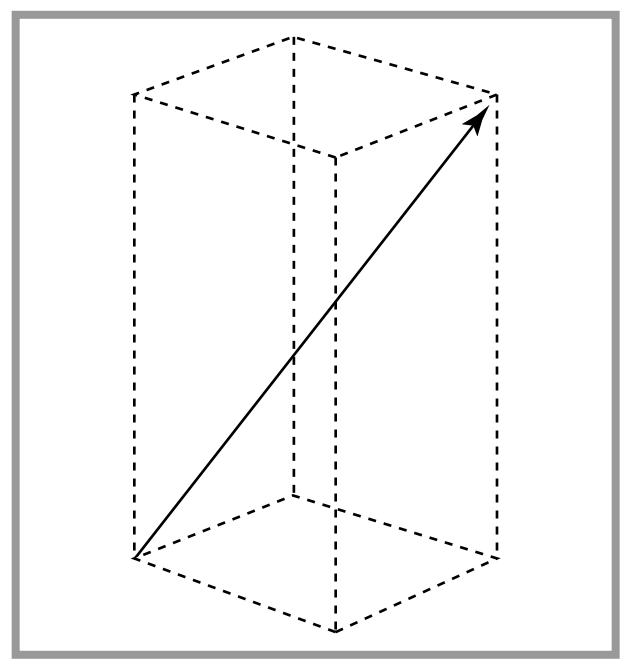


Figure 3.

Air traffic controllers use vectors (orders to fly a certain compass heading for a certain distance or time) to sequence aircraft during takeoff and landing. DAVID LAWRENCE/CORBIS.

stands for a force, then its magnitude gives the strength of that force. If a vector stands for an object's motion, then its magnitude gives how fast the object is moving.

The magnitude of a vector can be found by drawing it or building a model of it and measuring its length with a ruler, but it is easier to find the magnitude mathematically. To find the length of a vector that is known in numerical form, like the three-dimensional vector [4 4 8], first add the squares of all its parts or components. (The "square" of a number is that number multiplied by itself.) For example, to find the magnitude of [4 4 8], first calculate $(4 \times 4) + (4 \times 4) + (8 \times 8) = 96$. The second step is to find the square root of this sum. (The "square root" of a number is a second number that, when multiplied by itself, gives the first number. Square roots can be found using a calculator.) In this case, the sum is 96 and the square root of 96 is 9.798 because $9.798 \times 9.798 = 96$. Therefore, the magnitude of the vector [4 4 8] is 9.798. In the example where the arrow touching the two diagonally opposite corners of the box is described by the vector [4 4 8], the length of the arrow is 9.798 centimeters.

VECTOR ALGEBRA

Vectors can be added or multiplied according to the rules called "vector algebra." Addition and multiplication are different for vectors than for ordinary, "scalar" numbers.

Imagine that we want to add a second arrow or vector to the 5 cm vector shown earlier. This second vector is also 5 cm long, but points down rather than up. (See Figure 4.)

Let us call the first vector *A* and the second vector *B*. To perform vector addition of *A* and *B*, place them tip to tail and draw a third vector, *C*, from the base of the first arrow to the tip of the second, forming a triangle as depicted in Figure 5.

C is the vector sum of *A* and *B*. In vector algebra, the sum is written $A + B = C$. The vector *C* can also be found by adding the numerical versions of *A* and *B*. Here, *A* can be written [3 4] and *B* can be written [3 − 4]. *C* is found by adding the first dimension of *A* and *B* to give the first dimension of *C*, and adding the second dimension of *A* and *B* to give the second dimension of *C*: *C* = *[60]*.

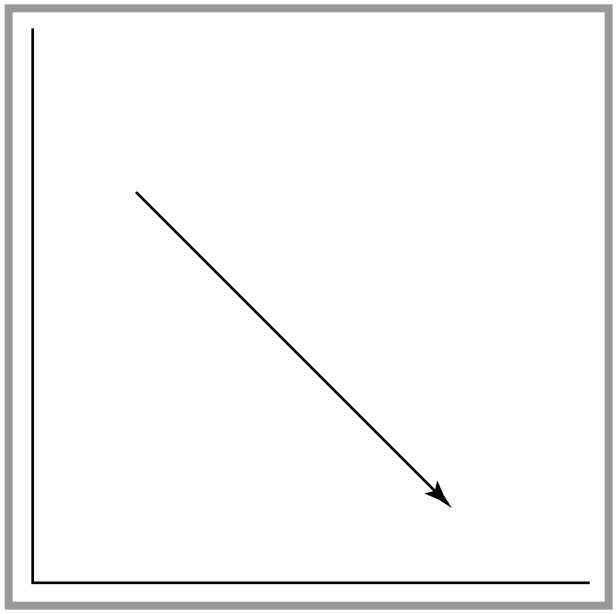


Figure 4.

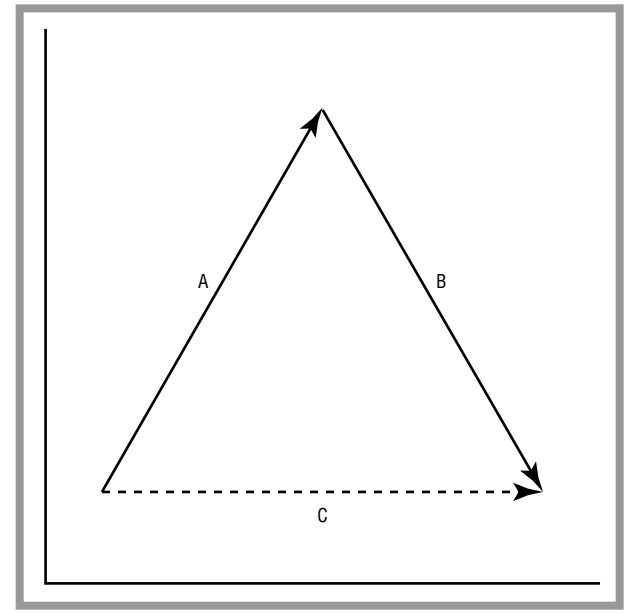


Figure 5.

Addition of vectors with more than two dimensions is done the same way.

Vector multiplication is more complicated. There is only one kind of multiplication for scalars—the $2 \times 2 = 4$ kind of multiplication—but there are two kinds of vector multiplication. The first, simpler kind is called the "inner product" and is written $A \cdot B = c$ (where A and B are vectors having the same number of dimensions and c is an ordinary number, a scalar). The inner product is found by adding up the products of the dimensions of A and B. For example, if $A = [3\ 4]$ and $B = [5\ 6]$, then $A \cdot B = (3 \times 5) + (4 \times 6) = 39$. The dot product of two vectors is a scalar, not a vector.

The second kind of vector multiplication is called the "vector product" or "cross product" of the two vectors. Some knowledge of trigonometry—the mathematical study of triangles—is needed to fully understand the vector product. The vector product of A and B is written $A \times B = C$, where A, B, and C are all vectors having the same number of dimensions. The magnitude and direction of C is calculated separately. The magnitude of C is given by multiplying the magnitude of vector A (written $|A|$), the magnitude of vector B (written $|B|$), and the sine of the angle between A and B (written τ): $A \times B = |A| \times |B| \times \sin\tau$. As for direction, C points at right angles to the plane containing the two vectors being multiplied. For example, if A and B are drawn on this piece of paper, then C sticks straight up out of the paper (or straight down into it, depending on the directions of A and B).

The elements of vectors may be functions rather than numbers. In the field known as "vector analysis," the methods of calculus are applied to such vectors.

VECTORS IN LINEAR ALGEBRA

We have already seen that it is easy to write a vector having more than two or three dimensions, such as [4 4 8 10 2]. This particular vector happens to have five dimensions. The rules for doing math with five-dimensional vectors are the same as for two- or three-dimensional vectors, but when vectors have more than three dimensions nobody tries to think of them as "arrows." In linear algebra, vectors with hundreds of dimensions may be used.

A Brief History of Discovery and Development

The idea of using vectors (arrows) to stand for velocities and the idea of adding two vectors by placing the arrows tip to tail to form a triangle were known to Greek thinkers over two thousand years ago. (This method of adding velocity vectors is also known as the "parallelogram of velocities" method, since the same answer can be found either by making the triangle or by making a parallelogram that has the two vectors as two of its edges.)

However, it was not until the 1600s that scientists began to handle many vector quantities, such as velocity,

force, momentum, and acceleration. In the 1700s and 1800s, science also began to deal more with other vectors in optics and electricity. This gradually forced the invention of better ways to handle vectors. In the late 1700s and 1800s, mathematicians struggling to deal with the question of "complex" numbers created two-dimensional vector algebras. The first important three-dimensional vector algebra was invented in the 1840s, at about the same time that matrix algebra was being created to handle matrices and vectors of higher dimensions. Since that time, the term "vector algebra" has come to refer mostly to the rules and symbols for handling vectors only (especially vectors of two, three, or four dimensions), while "matrix algebra" has come to refer to the rules and symbols for handling both vectors and matrices (rectangular arrays of numbers). Today, vectors are used through business, mathematics, and science.

Real-life Applications

3D COMPUTER GRAPHICS

3D ("three dimensional") computer graphics is the art or science of creating an imaginary world of spaces and objects inside a computer using numbers, then producing flat images from that imaginary world that can be shown on a screen or printed on paper. Popular movies such as *Shrek* and *The Incredibles* are produced using 3D computer graphics. The animators who make these movies rely on vectors at every stage. Vectors are used in 3D computer graphics to stand for the locations of points in the imaginary space, the edges of objects, the way objects are moving (velocities), and the way motions are changing (accelerations). Vector algebra is also used to create two-dimensional (flat) images of the computer's imaginary 3D world as seen from any angle desired. This involves the use of the dot product ($A \cdot B = C$, where A, B, and C are vectors) to find the "projections" of vectors defining object edges in the 3D world—that is, what all the vectors defining the edges of an object in the 3D world look like when seen from a certain angle. Some computer graphics programs also allow for realistic physics, that is, for "objects" in the imaginary world to fall or respond to pushes as if they were physically real. This, too, requires the use of vector algebra, since the motions of objects in response to forces are calculated using vectors.

DRAG RACING

To design a race car, engineers must take into account that a car is not a rigid block: it has a suspension system that allows its wheels to move. A car therefore changes overall shape temporarily whenever a force is applied to it, like acceleration, braking, or turning. All these forces must be treated as vectors.

When a car accelerates, several things happen at once. First, assuming that the car is rear-wheel drive, the rear tires push against the pavement. If they push too hard, they start to slip or spin, which is fine if the driver just wants to lay down rubber, but bad if he wants to win a race. The second thing that happens is that the force of the car's weight (a vector that points straight down, toward Earth's center) and the force accelerating the car (a vector pointing forward along the road) add to a total force vector that does not point straight down. Nor does it point straight through the car's center of mass. This effect is called "weight shift." Weight shift causes the car to "lift" in the front and "squat" in the rear, as if weights had been piled on the rear of the car. Because of weight shift, the front tires press on the road with less force and the rear tires press on the road with more force. This is good, up to a point, because more pressure on the back tires means that the car can accelerate faster without starting to slip. It is bad, however, if so much force is taken off the front tires that they rear right up off the road, which makes it impossible for the driver to steer. A drag racing car, therefore, must be designed using vector analysis so that when it is accelerating as hard as it can, the front wheels remain on the road—barely.

LAND MINE DETECTION

According to the United Nations, there are 60 to 100 million land mines buried around the world. These kill about 10,000 people a year—most of them civilians, not soldiers fighting wars—and maim about twice as many. Detecting landmines is therefore an urgent problem.

One method is to use radar in small, handheld units that are swept back and forth over the ground. The radar unit sends burst of radio waves down into the ground. Since landmines are made of metal and are usually round in shape, they reflect the radar differently than rocks, dirt, or differently-shaped metal objects such as wires and pipes. Still, it is not easy, even using a computer to analyze the radar reflections, to tell whether the radar is seeing a landmine or just "clutter," by which mine detection experts mean ground containing anything but land mines.

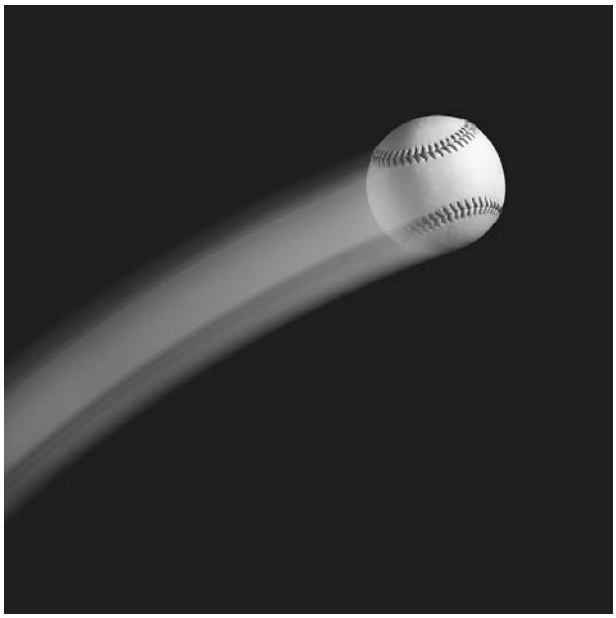
Vector analysis is being used to help make the detection of land mines more reliable. First, each signal bounced back from the ground is turned into a vector (an ordered series of numbers). This is done by finding the radar signal's "spectrum," a set of numbers that says how much of the signal's energy consists of vibrations at

different rates of speed: a signal with many vast vibrations will have energy at the high end of its spectrum, a signal with slow vibrations will have energy at the low end of its spectrum, and so on. Each radio echo from the ground therefore gives a vector.

The goal is to decide whether each vector consists of an echo from a land mine, or not. If the vectors were two-dimensional, then each could be plotted as an arrow on a piece of paper. Or, more simply, the tip of the arrow alone could be plotted, as a dot. The vectors from landmines are not always the same, so if many are plotted the dots will not be located all at one place, but in a fuzzy, rounded area called an "ellipse" (even though its outline it may not have the exact shape of a true geometric ellipse). The computer's job is to figure out whether any given vector is inside the ellipse, which would show a landmine was present. In practice, the vectors used in landmine detection may have from 40 to 128 dimensions, rather than two, but researchers still refer to the "ellipse" or "ellipsoid" (egg-shaped volume) in which the land mines are to be found.

SPORTS INJURIES

A part of the body often injured in sports is the ACL (anterior cruciate ligament). This is a tough, ribbon-like structure that helps keep the top part of the knee joint from slipping forwards and back (other ligaments keep the knee from slipping from side to side). ACL injuries are more common in sports that involve "cutting"—changing direction suddenly while running.



A baseball in motion. Its velocity at any instant is a three-dimensional vector. W. CODY/CORBIS.

Key Terms

Scalar: A directionless quantity with magnitude (e.g., speed as opposed to velocity which is a vector with both magnitude and direction).

Vector: A quantity consisting of magnitude and direction, usually represented by an arrow whose length represents the magnitude and whose orientation in space represents the direction.

Movement technique refers to how an athlete moves her or his body when making sports maneuvers. To study movement techniques, researchers begin by making measurements of male and female athletes in motion. They do so by attaching bright dots to various points on the athletes' bodies and then taking three-dimensional movies of them cutting, pivoting, jumping, and making other motions. They also place "force platforms" under the athletes' feet, flat plates that measure the strength and direction of the force vectors exerted by the athletes' feet against the ground.

These methods only produce raw data—lots of numbers, but no explanations. To understand what the raw data really mean, researchers must use many mathematical tools. One is vector analysis. Vectors are used to calculate the forces acting on the skeleton. They are also used to calculate the angle at which the knee is bent when the athlete's foot hits the ground (which is important to understanding how the knee is stressed). First, three-dimensional vectors standing for the length and position of the upper and lower parts of the leg are calculated from the video. Then the dot product of these two vectors is calculated by the method described earlier in this article. Since the dot product of two vectors depends on the angle between them, finding the dot product (plus one more mathematical step using trigonometry) gives us the angle.

Studies using these methods have shown that the knee is more vulnerable to ACL injury when it is less bent.

Where to Learn More

Books

Barnett, Raymond A., and John N. Fujii. *Vectors*. New York: John Wiley & Sons, 1963.

Crowe, Michael J. *A History of Vector Analysis*. Notre Dame, IN: University of Notre Dame Press, 1967.

Tallack, J.C. *Introduction to Vector Analysis*. Cambridge, UK: Cambridge University Press, 1970.

Periodicals

Slauterbeck, James. "Gender differences among sagittal plane knee kinematic and ground reaction force characteristics during a rapid sprint and cut maneuver." *Research Quarterly for Exercise and Sport*. Vol. 75, No. 1 (2004): 31–38.

Web sites

Olive, Jenny. "Working With Vectors." September 2003. <http://www.netcomuk.co.uk/~jenolive/homevec.html> (March 1, 2005).

Roal, Jim. "Automobile physics." AllFordMustangs.com. July 2003. <http://www.allfordmustangs.com/Detailed/29.shtml> (March 7, 2005).

"Vector Math for 3D Computer Graphics." Central Connecticut State University, Computer Science Department. July 2003. <http://chortle.ccsu.ctstateu.edu/VectorLessons/vectorIndex.html> (March 1, 2005).

Volume

Overview

An object's volume describes the amount of space it contains. Calculations and measurements of volume are used in medicine, architecture, science, construction, and business. Gasses and liquids such as propane, gasoline, and water are sold by volume, as are many groceries and construction materials.

Fundamental Mathematical Concepts and Terms

UNITS OF VOLUME

Volume is measured in units based on length: cubic feet, cubic meters, cubic miles, and so on. A cubic meter, for instance, is the amount of volume inside a box 1 meter (m) tall, 1 m wide, and 1 m deep. Such a box is a 1-meter cube, so this much volume is said to be one "cubic" meter. An object doesn't have to be a cube to contain a cubic meter: one cubic meter is also the space inside a sphere 1.24 meters across.

Cubic units are written by using exponent notation: that is, 1 cubic meter is written "1 m^3." This is why raising any number to the third power—that is, multiplying it by itself three times, as in $2^3 = 2 \times 2 \times 2$—is called "cubing" the number.

VOLUME OF A BOX

There are standard formulas for calculating the volumes of simple shapes. The simplest and most commonly used of these is the formula for the volume of a box. (By "box," we mean a solid with rectangular sides whose edges meet at right angles—what the language of geometry also calls a "cuboid," "right prism," or "rectangular parallelepiped.") To find the volume of a box, first measure the lengths of its edges. If the box is L centimeters (cm) long, W cm wide, and H cm high, then its volume, V, is given by the formula $V = L \text{ cm} \times W \text{ cm} \times H \text{ cm}$. This can be written more shortly as $V = LWH \text{ cm}^3$.

The units of length used do not make any difference to the formula for volume: inches or feet will do just as well as centimeters. For example, a room that is 20 feet (ft) long, 10 ft wide, and 12 feet high has volume $V = 20 \text{ ft} \times 10 \text{ ft} \times 12 \text{ ft} = 2{,}400 \text{ ft}^3$ (cubic feet).

VOLUMES OF COMMON SOLIDS

There are standard formulas for finding the volumes of other simple solids, too. Figure 1 shows some of these formulas.

Solid Shape	Dimensions	Formula for Volume
box	Length L, width W, height H	$V = LWH$
cube	Length = width= height = L	$V = L^3$
sphere	radius R	$V = \frac{4}{3}\pi R^3$
cylinder	radius R, height H	$V = \pi R^2 H$
cone	Base radius R, height H	$V = \frac{1}{3}\pi R^2 H$
pyramid	base area A, height H	$V = \frac{1}{3}AH$
torus (doughnut)	distance from center of torus to center of tube D, radius of tube R	$V = 2DR^2$

Figure 1: Standard formula to calculate volume.

In all these formulas, three measures of length are multiplied—not added. This means that whenever an object is made larger without changing its shape, its volume increases faster than its size as measured using a ruler or tape measure. For example, a sphere 4 m across (a sphere with a radius of 2 m) has a volume of $V = 4/3\ \pi 2^3 = 33.5\ \text{m}^3$, whereas a sphere that is twice as wide (radius of 4 m) has a volume of $V = 4/3\ \pi\ 4^3 = 268.1\ \text{m}^3$. Doubling the radius does not double the volume, but makes it 8 times larger. In general, since the radius is cubed in calculating the volume, we say that a sphere's volume "increases in proportion to" or "goes as" the cube of its radius. This is true for objects of all shapes, not just spheres: Increasing the size of an object without changing its shape makes its volume grow in proportion to the cube (third power) of the size increase.

The formula for an object's volume can be compared to the formula for its area. The area of a sphere of radius R, for example, is $A = 4\ \pi\ R^2$. The radius appears only as a squared term (R^2) in this formula, whereas in the volume formula it appears as a cubed term (R^3). Dividing the volume formula by the area formula yields an interesting and useful result:

$$\frac{V}{A} = \frac{\frac{4}{3}\pi R^3}{4\pi R^2}$$

Crossing out terms that are the same on the top and bottom of the fraction, we have

$$\frac{V}{A} = \frac{1}{3}R$$

which, if we multiply both sides by A, becomes

$$V = \frac{AR}{3}$$

This means that when we increase the radius R of a sphere, area and volume both increase, but volume increases by the increased area times $R/3$. Volume increases faster than area. This fact has important consequences for real-world objects. For example, how easily an animal can cool itself depends on its surface area, because its surface is the only place it can give heat away to the air; but how much heat an animal produces depends on its volume, because all the cubic inches of flesh it contains must burn calories to stay alive. Therefore, the larger an animal gets (while keeping the same shape), the fewer square inches of heat-radiating skin it has per pound: its volume increases faster than its area. A large animal in a cold climate should, therefore, have an easier time staying warm. And in fact, animals in the far North tend to be bigger than their close relatives farther south. Polar bears, for example, are the world's largest bears. They have evolved to large size because it is easier for them to stay warm. On the other hand, a large animal in a hot climate has a harder time staying cool. This is why elephants have big ears: the ears have tremendous surface area, and help the elephant stay cool.

Volume can be described in terms of an amount of the space an object assumes, such as water in a bucket.
ROYALTY-FREE/CORBIS.

A Brief History of Discovery and Development

Weights, lengths, areas, and volumes were the earliest measurements made by humankind. Not only are they easier to measure than other physical quantities, like velocity and temperature, but they have an immediate money value. Measuring lengths, builders can build more complex structures, such as temples; measuring area, landowners can know how much land, exactly, they are buying and selling; measuring volume, traders can tell how much grain a basket holds, or how much water a cistern (holding tank) holds. Therefore it is no surprise to find that the Egyptians, Sumerians, Greeks, and ancient Chinese all knew the concept of volume and knew many of the standard equations for calculating it. In 250 B.C. (over 2,200 years ago), the Greek mathematician Archimedes wrote down formulas for the volume of a sphere and cylinder. In approximately 100 B.C., the Chinese had formulas for the volumes of cubes, cuboids, prisms, spheres, cylinders, and other shapes (using, like the Greeks, approximate values for π ranging from rough to excellent).

Such formulas are useful but do not give any way of exactly calculating the volume of a shape whose surface is not described by flat planes or by circles (as are the curved sides of a cylinder, or the surface of a sphere). New progress in the calculation of volumes had to wait almost 2,000 years, until the invention of the branch of mathematics known as calculus in the 1600s. One of the two basic mathematical operations of calculus is called "integration." Integration, as it was first invented, allowed mathematicians to exactly calculate the area under any mathematically defined curve or any part of such a curve; it was soon discovered, however, that integration was not restricted to flat surfaces and areas. It could be generalized to three dimensions—that is, to ordinary space. It had now become possible to calculate exactly the volumes of complexly-shaped objects, as long as their surfaces could be described by mathematical equations.

The next great revolution in volume calculation came with computers. Since computers can add many numbers very quickly, they have made it possible to calculate areas and volumes for complex shapes even when the shapes cannot be described by nice, neat mathematical equations. Today, the calculation of volumes of simple shapes is still routine in many fields, but the use of calculus and computers for complex shapes such as airplane wings and the human brain is increasingly common.

Real-life Applications

PRICING

Volume is closely related to density, which is how much a given volume of a substance weighs. For instance, the density of gold is 19.3 grams per cubic centimeter, that is, one cubic centimeter of gold weighs 19.3 grams, which is 19.3 times as much as one cubic centimeter of water. Silver, platinum, and other metals all have different densities. This fact is used by some jewelry makers to decide how much to charge for their jewelry.

Different metals not only have different densities, they have different costs: at a 2005 price of about $850 per ounce, for example, platinum cost about twice as much as gold. So when a jewelry maker uses a blend of gold and platinum in a piece of jewelry, they need to know exactly how much of each they have used in order to know how much to charge for the piece. Now, a blend of two metals (called an alloy) has a density that is somewhere between the densities of the two original metals. Therefore, determining the average density of a piece (say, a ring) will tell a manufacturer how much gold and platinum it contains, regardless of how complicated the piece is. Volume and weight together are used to determine density. The finished piece is suspended in water by a thread. Any object submerged in water experiences an upward force that depends only on the volume of water the object displaces. Therefore, by weighing the piece of jewelry as it hangs in water, and comparing that weight to its weight out of water, the jeweler can measure exactly what weight of water it displaces. Since the density of water is known (1 gram per cubic centimeter), this water weight tells the jeweler the exact volume of the piece. Finally, knowing both the volume of the piece and its weight, the jeweler can calculate its density by the equation density

= weight/volume. The jewelry maker's wholesale price will be determined partly by this calculation, and so will the retail price in the store.

MEDICAL APPLICATIONS

In medicine, volume measurements are used to characterize brain damage, lung function, sexual maturity, anemia, body fat percentage, and many other aspects of health. A few of these uses of volume are described below.

Brain Damage from Alcohol Using modern medical imaging technologies such as magnetic resonance imaging (MRI), doctors can take three-dimensional digital pictures of organs inside the body, including the brain. Computers can then measure the volumes of different parts of the brain from these digital pictures, using geometry and calculus to calculate volumes from raw image data.

MRI volume studies show that many parts of the brain shrink over time in people who are addicted to alcohol. The frontal lobes—the wrinkled part of the brain surface that is just behind the forehead—are strongly affected. It is this part of the brain that we use for reasoning, making judgments, and problem solving. But other parts of the brain shrink, too, including structures involved in memory and muscular coordination. Alcoholics who stop drinking may regain some of the lost brain volume, but not all. MRI studies also show that male and female alcoholics lose the same amount of brain volume, even though women alcoholics tend to drink much less. Doctors conclude from this that women are probably even more vulnerable to brain damage from alcohol than are men.

Diagnosing Disease Almost half of Americans alive today who live to be more than 85 years old will suffer eventually from Alzheimer's disease. Alzheimer's disease is a loss of brain function. In its early stages, its victims sometimes have trouble remembering the names for common objects, or how they got somewhere, or where they parked their car; in its late stages, they may become incurably angry or distressed, forget their own names, and forget who other people are. Doctors are trying understand the causes of Alzheimer's disease and develop treatments for it. All agree that preventing the brain damage of Alzheimer's—starting treatment in the early stages—is likely to be much more effective than trying to treat the late stages. But how can Alzheimer's be detected before it is already damaging the mental powers of the victim?

Recent research has shown that the part of the brain called the hippocampus, which is a small area of the brain located in the temporal lobe (just below the ear), is the first part of the brain to be damaged by Alzheimer's. The hippocampus helps the brain store memories, which is why forgetting is one of Alzheimer's first symptoms. But instead of waiting for memory to fail badly, doctors can measure the volume of the hippocampus using MRI. A shrinking hippocampus can be observed at least 4 years before Alzheimer's disease is bad enough to diagnose from memory loss alone.

Pollution's Effects on Teenagers Polychlorinated aromatic hydrocarbons (PCAHs) are a type of toxic chemical that is produced by bleaching paper to make it white, improper garbage incineration, and the manufacture of pesticides (bug-killing chemicals). These chemicals, which are present almost everywhere today, get into the human body when we eat and drink. In 2002 scientists in Belgium studied the effects of PCAHs on the sexual maturation of boys and girls living in a polluted suburb. They compared how early boys and girls in the polluted suburb went through puberty (grew to sexual maturity) compared to children in cleaner areas. They found that high levels of PCAH-related chemicals in the blood significantly increased the chances of both boys and girls of having delayed sexual maturity. Once again, volume measurements proved useful in assessing health. The researchers estimated the volume of the testicles as a way of measuring sexual maturity in boys, while they assessed sexual maturity in girls by noting breast development. This study, and others, show that some pollutants can injure human health and development even in very low concentrations. Testicular volume measurements are also used in diagnosing infertility in men.

Body Fat Doctors speak of "body composition" to refer to how much of a person's body consists of fat, muscle, and bone, and where the fat and muscle are located on the body. Measuring body composition is important to monitoring the effects of diet and exercise programs and tracking the progress of some diseases. Volume measurement is used to measure some aspects of body composition. For example, the overall density of the body can be used estimate what percentage of the body consists of fat. Measuring body density requires the measurement of the body's weight—which can be done easily, using a scale—and two volumes.

The first volume needed is the volume of the body as a whole. Since the body is not made of simple shapes like cubes and cylinders, its volume cannot be found by taking a few measurements and using standard geometric formulas. Instead, its volume must be measured by submerging it in water. The body's overall volume can then be found by measuring how much the water level rises or, alternatively, by weighing the body while it is underwater to see how much water it has displaced. (Underwater weighing is the same method used to measure the density

of jewelry containing mixed metals, as described earlier in this article.) The body's overall volume is equal to the water displaced.

However, doctors want to know the weight of the solid part of the body; the air in the lungs does not count. And even when a person has pushed all the air they can out of their lungs, there is still some left, the "residual lung volume." Residual lung volume must therefore also be measured, as well as overall body volume. This is done using special machines that measure how much gas remains in the lungs when the person exhales. The body's true, solid volume is approximately calculated by subtracting the residual lung volume from the body's water displacement volume.

Dividing the body's weight by its true, non-air volume gives its density. This is used to estimate body fat percentage by a standard mathematical formula.

BUILDING AND ARCHITECTURE

Many building materials are purchased by area or volume. Area-purchased materials include flooring, siding, roofing, wallpaper, and paint. Volume-purchased materials include concrete for pouring foundations and other structures, sand or crushed rock, and grout (a kind of thin cement used to fill up masonry joints). All these materials are ordered by units of the cubic yard. (One cubic yard equals about .765 cubic meters.) In practice, simple volume formulas for boxes and cylinders are used to calculate how many cubic yards of cement must be ordered to build simple structures like housing foundations. A simple foundation, shaped like a box without a top, can be broken into three slab-shaped boxes, namely the four walls and the floor. Multiplying the length by the width by the thickness of each of these slabs gives a volume: the sum of these volumes is the cubic yardage that the cement truck must deliver. For concrete columns, the formula for the volume of a cylinder is used. For complex structures with curving shapes, a computer uses calculus-based methods to calculate volumes based on digital blueprints for the structure.

The same principle is used in designing machine parts. It is necessary to know the volume of a machine part while it is still just a drawing in order to know what its weight will be: its weight must be known to calculate how much it will weigh, and (if it is a moving part) how much force it will exert on other parts when it moves. For parts that are not too complicated in shape, the volume of the piece is calculated as a sum of volumes of simple elements: box, cylinder, cone, and the like. Computers take over when it is necessary to calculate the volumes of pieces with strange or curvy shapes.

COMPRESSION RATIOS IN ENGINES

Internal combustion engines are engines that burn mixtures of fuel and air inside cylinders. Almost all engines that drive cars and trucks are of this type. In an internal combustion engine, the source of power is the cylinder: a round, hollow shaft sealed at one end and with a plug of metal (the piston) that can slide back and forth inside the shaft. When the piston is withdrawn as far as it will go, the cylinder contains the maximum volume of air that it can hold: when the piston is pushed in as far as it will go, the cylinder contains the minimum volume of air. To generate power, the cylinder is filled with air at its maximum volume. Then the piston is pushed along the cylinder to compress the air. This makes the air hotter, according to the well-known Ideal Gas Law of basic physics—just how hot depends on how small the minimum volume is. Fuel is squirted into the small, hot volume of air inside the cylinder. The mixture of fuel and air is then ignited (either by sheer heat of compression, as in a diesel engine, or by a spark plug, as in a regular engine) and the expanding gas from the miniature explosion pushes the piston back out of the cylinder. The ratio of the cylinder's largest volume to its smallest is the "compression ratio" of the engine: a typical compression ratio would be about 10 to 1. Engines with high compression ratios tend to burn hotter, and therefore more efficiently. They are also more powerful. Unfortunately, there is a dilemma: burning very hot (high compression ratio) allows the nitrogen in air to combine with the oxygen, forming the pollutant nitrogen oxide; burning relatively cool (low compression ratio) allows the carbon in the fuel to combine only partly with the oxygen in the air, forming the pollutant carbon monoxide (rather than the non-poisonous greenhouse gas carbon dioxide).

GLOWING BUBBLES: SONOLUMINESCENCE

When small atoms come together to make a single heavier atom, energy is released. This process is called "fusion" because in it, two atoms fuse into one. All stars, including the Sun, get their energy from fusion. Some nuclear weapons are also based on fusion. But fusion is difficult to control on Earth, because atoms only fuse under extreme heat. If fusion could be controlled, rather than exploding as a bomb, it could be used to generate electricity. Many billions of dollars have been spent on trying to figure out how to make atoms trapped inside magnetic fields fuse—so far without success.

Yet there is a new possibility. Some reputable scientists claim that they can produce fusion using nothing more expensive or exotic than a jar full of room-temperature

liquid bombarded by sound waves. This claim—which has not yet been tested by other researchers—is related to the effect called "sonoluminescence," which means "sound-light." Sonoluminescence depends on changes in volume of bubbles in liquid. Under certain conditions, tiny bubbles form and disappear in any liquid that is squeezed and stretched by strong sound waves; when the bubbles collapse, they can emit flashes of light. This happens as follows: Pummeled by high-frequency sound waves, a bubble forms and expands. When the bubble collapses, its radius decreases very rapidly as its surface moves inward at several times the speed of sound. Because the volume of a sphere is proportional to the cube (third power) of its radius, when a bubble's radius decreases to 1/10 of its starting value, its volume decreases to $(1/10)^3 = 1/1{,}000$ of its starting value. (These are typical figures for the collapse of a sonoluminescence bubble.) This decrease in volume squeezes the gas inside the bubble, and, according to laws of physics, when a gas is squeezed its temperature goes up. Also, the compression happens very quickly—too quickly for much heat to escape from the bubble. Therefore, the bubble's rapid shrinkage causes a fast rise in temperature inside the bubble. The temperature has been shown to rise to tens of thousands of degrees, and may reach over two hundred thousand degrees. Such heat rivals that at the heart of the Sun and makes the gas in the bubble glow. It may also do something else: in 2002 scientists at Oak Ridge National Laboratory claimed to have detected neutrons flying out of a beaker of fluid in which sonoluminescence was occurring. Neutrons would be a sign that fusion was occurring. If it is, then there is a close resemblance between bubble fusion and the diesel engines found in trucks: both devices work by rapidly decreasing the volume of a gas in order to heat it to the point where energy is released. In a diesel engine, the energy is released by a chemical reaction. In a fusion bubble, it would be released by a nuclear reaction.

As of 2005, the reality of bubble fusion had been neither proved nor disproved. If it is proved, it might eventually mean that producing electricity from fusion could be done more cheaply than scientists had ever before dreamed. Describing changes in bubble volume mathematically is basic to all attempts to understand and control sonoluminescence and bubble fusion.

SEA LEVEL CHANGES

One of the potential threats to human well-being from possible global climate change is the rising of sea levels. The International Panel on Climate Change predicts that ocean levels will rise by 3.5 inches to 34.5 inches (about 9 to 88 centimeters) by the year 2100, with a best guess of 1.6 ft (about 50 centimeters) with the ocean continuing to rise. Hundreds of millions of people live near sea level worldwide, and their homes might be flooded or at greater risk from flooding during storms. Also, many small island nations might be completely flooded.

Sea level rises when the volume of water in the ocean increases. There are two ways in which a warmer Earth causes the volume of water in the ocean to increase. First, there is the melting of ice. Ice exists on Earth mostly in the form of glaciers perched on mountain ranges and the ice caps at the north and south poles. Second, there is the volume increase of water as it gets warmer. Like most substances, water expands as it gets warmer: a cubic centimeter of seawater gains about .00021 cubic centimeters of volume if it is made 1 degree Centigrade warmer. Therefore, the oceans get bigger just by getting warmer. In fact, the International Panel on Climate Change predicts that most of the sea-level rise that will occur in this century will be caused by water expansion, rather than by ice melting and increasing the mass of the sea. Calculations of the volume of water that will be added to the ocean by melting glaciers and icecaps and by thermal expansion are at the heart of predicting the effects of global warming on sea levels.

WHY THERMOMETERS WORK

The fact that liquids expand as they get warmer (until they start to boil) is used to measure temperature in old-fashioned mercury or colored-alcohol thermometers. Geometry is used to amplify or multiply the expansion effect: a thin cylinder connected attached to a sphere (the "bulb"). The bulb is full of liquid. If the radius of the thermometer bulb is r_B, then its volume (V_B, for "volume, bulb") is given by the standard volume formula for a sphere as

$$V_B = \frac{4}{3}\pi r_B^3$$

If the cylinder's radius is r_C, then the volume of liquid in the cylinder (V_C, for "volume, cylinder") is given by the standard volume formula for a cylinder as $V_C = \pi r_C^2 H$, where H is the height of the fluid in the cylinder. We read the temperature from a thermometer of this type by reading H from marks on the cylinder.

There is room in the cylinder for more liquid, but there is no room in the sphere, which is full. If the thermometer contains a liquid that has a "volume thermal coefficient" of $\alpha = .0001$, a cubic centimeter of the liquid will gain .0001 cubic centimeters of volume if it is warmed by 1 degree Centigrade. Say that the thermometer starts

out with no fluid in the cylinder and the bulb perfectly full. Then the temperature of the thermometer goes up by 1°C. This causes the volume of the fluid in the bulb, V_B before it is warmed, to increase by $.0001 V_B$. But this extra volume has nowhere to go in the bulb, which is full, so it goes up the cylinder. The amount of fluid in the cylinder is then $V_C = \pi r_C^2 H = .0001 V_B$. If we divide both sides of this equation by πr_C^2, we find that

$$H = \frac{.001 V_B}{\pi r_C^2}$$

Because V_B is on top of the fraction, making it bigger makes H bigger. That is, the bigger the bulb, the bigger the change in the height of the fluid in the cylinder when the temperature goes up. Since r_C is on the bottom of the fraction, making it smaller also makes H bigger. That is, the narrower the cylinder, the bigger the change in the height of the fluid in the cylinder when the temperature goes up. This is why thermometers have very narrow cylinders attached to fat bulbs—so it is easy to see how far the fluid goes up or down the cylinder when the temperature changes.

MISLEADING GRAPHICS

Many newspapers and magazines think that statistics are dull, and so they have the people who work in their graphics departments make them more visually appealing. For example, to illustrate money inflation (how a Euro or a dollar buys less every year), they will show you a picture of shrinking bill—a big bill, then a smaller bill below it, and a smaller below that, and so forth. Or, to illustrate the increasing price of oil, they will show you a picture of a row of oil barrels, each bigger than the last.

Such pictures can create a very false impression, because it is usually the lengths of the dollar bills or the oil barrels (or whatever the object is), not their areas or volumes, that matches the statistic the art is trying to communicate. So, to show the price of oil going up by 10%, a publication will often show two barrels, one 10% taller and wider than the other. But the equation for the volume of a barrel, which is a cylinder, is $V = \pi r^2 H$, where r is the radius of the barrel and H is its height. Increasing r or H by 10% is the same as multiplying it by 1.1, so increasing the dimensions of the barrel by 10% shows us a barrel whose volume is $V_{\text{bigger}} = \pi (1.1\ r)^2 (1.1)H$. If we multiply out the factors of 1.1, we find that $V_{\text{bigger}} = 1.331V$—that is, the volume of the larger barrel in the picture, the amount of oil it would contain, is not 10% larger but 33.1% larger. Because volume increases by the cube of the change in size, the larger the size change, the more misleading the picture.

Look carefully at any illustration that shows growing or shrinking two-dimensional or three-dimensional objects to illustrate one-dimensional data (plain old numbers that are getting larger or smaller). Does the artwork exaggerate?

SWIMMING POOL MAINTENANCE

Everyone who owns a swimming pool knows that they have to add chemicals to keep the water healthy for swimming. It's not enough to just dump in a bucket or two of aluminum sulfate or calcium hypochlorite, though—the dose has to be proportioned to the volume of water in the pool.

Some pools have simple, box-like shapes: their volume can be calculated using the standard formula for the volume of a box, volume equals length times width times height. A standard formula can also be used for a circular pool with a flat volume, which is simply a cylinder of water. Many pools have more complex shapes, though, and even a rectangular pool often has a deep end and a shallow end. The deep and shallow ends may be flat, with a step between, or the bottom of the pool may slope. Some pools are elliptical (shaped like a stretched circle), and an elliptical pool may also have a sloping bottom.

To calculate the correct chemical dose for a swimming pool, it is necessary, then, to take some measurements. A pool with a complex shape has to be divided into sections with simpler shapes, and the volumes of the separate pieces calculated and added up. More complex formulas are needed for, say, the volume of an elliptical pool with a sloping bottom; calculus is needed to find these formulas. Fortunately for the owners of complexly shaped pools, volume-calculation computer software exists that will calculate a pool's volume given the basic measurements of its shape. For an elliptical pool with a sloping bottom, you would need to measure the length of the pool, the width of the pool, the maximum depth, and the minimum depth.

BIOMETRIC MEASUREMENTS

On average, men's brains tend to be larger than women's, occupying more volume and weighing more. Before the invention of modern medical imaging machines like CAT (computerized axial tomography) scanners, brain volumes were measured by measuring the volumes of men's and women's skulls after they were dead. Beads, seeds, or ball bearings were poured into the empty skull to see how much the skull would hold, then they were weighed. More beads, seeds, or bearings meant more

brain volume. Today, brain volume can be measured in living people using computer software that uses three-dimensional medical scans of the brain to count how many cubic centimeters of volume the brain occupies.

But the fact that men, on average, have slightly larger brains (about 10% larger) does not mean that men are smarter than women. To begin with, a bigger brain does not mean a more intelligent mind, and there is great individual variation among people of both sexes. Some famous scholars have been found, after death, to have brains only half the size of other scholars. People of famous intelligence, like Einstein, usually do not have larger-than-average brain volume. Second, about half of the average size difference is accounted for by the fact that men tend to be larger than women. Brain size goes, on average, with body size: taller, more muscular men tend to have larger brains than smaller, less muscular men. Elephants and whales have larger brains several times larger than those of human beings, but are not more intelligent. To some extent, therefore, men have larger brains only because their bodies are larger, too.

In the nineteenth and early twentieth centuries, brain-volume measurements were used to justify laws that allowed only men to vote and hold some other legal rights. This is a classic case of accurate measurements being interpreted in a completely misleading way.

RUNOFF

Runoff is water from rain or melting snow that runs off the ground into streams and rivers instead of soaking into the ground. Scientists and engineers who study flood control, sewage management, generating electricity from rivers, shipping goods on rivers, or recreation on rivers make determinations of water volume to estimate supply. To make an educated guess, they initially estimate the volume of water that will be added by snowmelt and rainfall during a given period of time. This indicates how much water will arrive, and when and how fast, in various rivers or lakes.

Hydrogeologists and weather scientists use complex mathematical equations, satellite data, soil-test data, and computer programs to predict runoff volumes. Some of the factors that they must take into account include rain amount, intensity, duration, and location; soil type and wetness; snowpack depth and location; temperature and sunshine; time of year; ground slope; and the type and health of the vegetation covering the ground. All this information goes into a mathematical model of the stream, lake, or reservoir basin into which the water is draining. Given the exact shape of the basin receiving the water, water volume can be translated into water depth. In some places, water can be drained from reservoirs to make room for the volume of water that has been forecast to flow from higher ground, thus preventing floods.

Where to Learn More

Books

Tufte, Edward R. *The Visual Display of Quantitative Information.* Cheshire, CT: Graphics Press, 2001.

Web sites

"Causes of Sea Level Rise." Columbia University, 2005. <http://www.columbia.edu/~epg40/elissa/webpages/Causes_of_Sea_Level_Rise.html> (April 4, 2005).

"Making a River Forecast." US National Weather Service, Sep. 21, 2004. <http://www.srh.noaa.gov/wgrfc/resources/making_forecast.html> (April 6, 2005).

"Volume." Mathworld. 2005. <http://mathworld.wolfram.com/Volume.html> (April 4, 2005).

Word Problems

Overview

The ability to communicate and the development of language have paralleled the progression in society of mathematical and scientific developments. Humans think and imagine in language and pictures, so it is hardly surprising that much of mathematics deals with the translation from words to expressions. The word translate can be used because many people view math as a language in its own right. After all, it has its own rules of grammar and layout. It should also be perfectly logical.

It is often observed that a good mathematician is one who can translate complicated real-life situations into logical mathematical sentences that can then be solved.

Fundamental Mathematical Concepts and Terms

There are two distinct types of word problems, both relevant to today's world. First, there is the statement believed to true. Mathematics can often be used to establish the validity of the statement. This proposition is often called a hypothesis. Often a written statement can be proven to hold true without exception. These ideas branch out into a large mathematical area called proof. There are many different ways of proving things. These proofs can often have tremendous impact on the real world because people can the use these ideas completely and confidently.

Second, there is the word problem, to which the solution happens to involve mathematics. Mathematical modeling is considered to be the process of turning real-life problems into the more abstract and rigorous language of mathematics. It generally involves assumptions and simplifications required to express the complex situation as one that can be solved.

These solutions are then compared to the actual readings or observations. Alterations are then made to the model to try to achieve a more realistic solution. These alterations are often referred to as refinements. This process of solving, comparing, and refining is called the modeling process. It is used to solve many of the problems in the real world. It is used because it is often impossible to exactly model the frequently immeasurable possibilities in real life. Simplifications often lead to a realistic and useable model.

Diagrams are also used to simplify situations. The key elements can be marked and these are then used within the model. One of the key facts that should be

considered is that a diagram will help simplify even the most complex of problems.

A Brief History of Discovery and Development

It is frequently the case that the person involved as a manager behind a job will have the ideas but not the mathematical ability to solve the problem. It is for this reason that mathematics, whether through mathematicians, engineers, scientists, or statisticians, is thus employed.

Possibly one of the early cases of such an idea was the building of ancient monuments some of which, it is now believed, tell time and measure the passing of seasons. The most famous example includes the building of the pyramids. The pharaohs, wanting to express their might and wealth, commanded the building of these tombs without the slightest idea of the mathematics behind them. It was the engineers who set to work, translating the request into achievable, long-lasting designs.

As the years have progressed, so the requests and subsequent designs have become and more detailed and complicated. War, however terrible, has forced great strides in our technologies. Requests for fighting machines have driven much of the mathematics behind flight, engines, and electronics. Progress in trade and finance has also forced people into solving problems involving money. Though these calculations generally use the four basic operators, (add, subtract, divide, and multiply), the ability to translate between statements and calculations is a highly sought after skill. The more complex finance has become, so the complexity of problems met in the real world has increased.

Perhaps the biggest driving force is the current emphasis towards efficiency. It is increasingly the case that the best solutions, often referred to as optimal solutions, are required. Today, only the very best will do.

Real-life Applications

TEACHERS

Teachers spend most of their time trying to construct real-life problems. It is widely believed that understanding the mathematics behind actual problems assists in grasping the more theoretical, fundamental, and abstract ideas that underpin mathematics. It also makes the subject more accessible, relevant, and interesting. Indeed, it is the application to real life that has driven many of the advancements in mathematics. The more abstract side of mathematics is a beautiful area, and application to the real world provides a stepping-stone into this complex and remarkable subject.

COMPUTER PROGRAMMING

Computers are built with an underlying logic behind them. This logic is used to then program software or games. The computer designer will have ideas about how to make the interface look and how to program the operating software to allow for a suitable user-friendly environment.

SOFTWARE DESIGN

The design of software goes through various processes. First, the creative department will come up with ideas for a suitable game. This will often be deduced through market research. The department will then pass on ideas to the programmers, who will translate the creative ideas into programming code. Programming code is an example of the use of mathematics. It follows a logical structure and obeys the many structures underlying mathematics.

CREATIVE DESIGN

The artistic idea behind animation, computer graphics, or a storyline will often be verbal. This then has to be turned into motion through the work of computer designers. Highly competent mathematicians will program these packages. The concepts behind three dimensions, perspective, etc. have to be converted into machine code. These are effectively strings of mathematical statements. They will use vast arrays (data storage) that are then manipulated.

INSURANCE

Insurance involves almost exclusively real-life situations. A client will provide a list of items that need to be insured against loss, and the insurance company will then try to offer an attractive premium that the client will be willing to pay to insure his items. The evaluating of such premiums can be a highly complex task. The people involved, who are often referred to as actuaries, need to simplify all the variables involved and work out the various probabilities. Not only do they want to encourage the client to pay the premium, they must also ensure that, on average, the company will not lose vast sums of money in event of a claim.

Actuaries evaluate what is often referred to as the expected monetary value of the situation. This is simply the expected financial outcome of a given financial situation.

They will often draw a simple tree diagram, upon which expected occurrences are labeled. They can then work out from this the best possible premium for the situation.

This allows solutions to such questions as, What is the best premium? How much should be charged? It also allows the consumer to evaluate the best deal being offered. Everyone, at some point in life, will be faced with the prospect of buying insurance. Every first-time driver will be expected to pay a premium that is much greater that experienced drivers.

CRYPTOGRAPHY

Cryptography is the ability to send encoded data that, in theory, will be unreadable without a key. Authorities need to be able to control and often intercept messages and then read them. In modern times, where terrorism is often referred to as a significant threat, it is essential to be able to understand what such groups are saying. By its very nature, cryptography deals problems involving words.

There are many different ways of coding data, yet an awareness of the different possibilities means that, with powerful computers, a piece of writing can be unscrambled in many different ways until the correct key is found. The ability to decode information can hinge on knowledge of the actual language used. However a coding is applied, the frequency of certain letters within the language can be used to try to decode simple situations. During World War II, decoding was often found to be difficult due to the placing of random letters into specific sections of the text, but the decoders generally prevailed.

MEDICINE AND CURES

Research in medicine is frequently concerned with questioning the benefits of drugs as well as assessing their possible side effects. It is an extremely difficult area to research, because people's lives are so heavily mixed into the equation. It is impossible to test all drugs on all people and record which ones work while recording the visible effects on the patients. So, how does a question such as "Does smoking cause cancer?" actually get solved mathematically?

These are questions involving causality. Namely, does smoking actually cause cancer? It is often the case that, even though there appears to be a direct link, it is either a fluke or a third variable is causing the apparent situation. To determine this, strict statistical tests need to be carried out using a control group, made up of people that have no link to the drug in question. Another group is then selected, who are given just the drug. These people would have to be selected randomly to reduce the chance of a third variable. The outcomes can then be compared and inferences drawn.

HYPOTHESIS TESTING

This is an important area of mathematics. It is equivalent to a court case, in which a party is only found guilty if the evidence is of sufficient nature. For instance, it is believed that playing computer games has caused a decrease in the number of people reading books. To prove this, the situation is set up extremely systematically. A null hypothesis is defined. This often states a simple belief that there has been no change: computers have not caused a decrease in literacy.

An alternative hypothesis is defined. This would be a statement indicating that there has been a change. In this case, computers have decreased literacy. A statement is then made indicating how much evidence is required to decide on the alternative hypothesis. This is called the significance. The statistician would then pick a random sample of people relevant to the survey. These would have to be drawn from the whole population. The statistician would then take a survey on reading and computer habits and compare this to data from the past. If the change (presuming a change) were to be sufficient, it would be stated that there existed enough evidence for the alternative hypothesis.

In hypothesis testing it is essential to define the significance before the test, otherwise the conclusion may be compromised.

ARCHAEOLOGY

Archaeology uses many mathematical ideas to analyze many different aspects, from dating individual objects to how the landscape has changed. These facts are then pieced together to provide an overall picture to help in understanding the past.

ENGINEERING

The conversion of ideas into safe and workable designs involves a lot of detailed mathematics. For instance, how does water arrive through the tap? The many different stages in the process would be separated and each part solved progressively. The whole system involves forces, which allow the water to flow around the system. This in turn puts pressure on the system; hence it needs to be strong enough and yet cheap enough to run. A single error in calculation along the way and the whole process would have to be thought through again at much

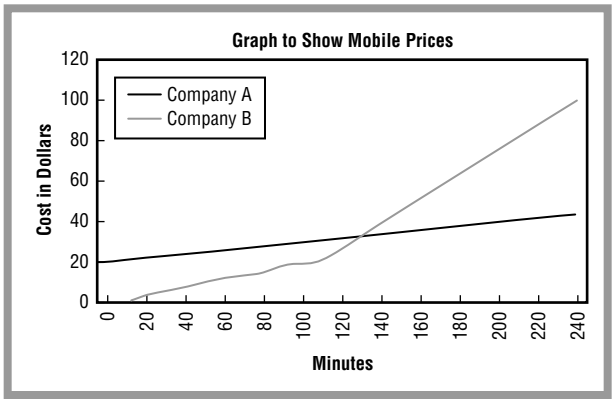


Figure 1.

expense. The sewerage and water system beneath any major city is a great engineering and mathematical feat.

COMPARISONS

Statements are often made concerning views on sports persons or other famous figures such as pop stars. Frequent allusions are made to the best ever sportsman or the most successful singer. Mathematics is used to solve such problems using the concept of averages. There are three main types of averages: mean, median, and mode, each having an exact meaning.

For example, a teacher has stated that Sam is better at math than David. This is because Sam averages 70, while David averages 65. Sam's scores were 40, 70, and 80; David's scores were 65, 65, and 100. It is perhaps immediately apparent that David has the better scores overall. When solving problems involving averages, it is also useful to indicate how spread out the data is. This indicates how consistent someone or the object in question is.

PERCENTAGES

Everyday, the consumer is confronted by billboards offering massive savings and bargain prices in an attempt by retailers to tempt the customer in. The customer must see a way around any potential pitfalls. For instance, if a store suggests that 40% of their competitors are worse than they are, the clever consumer would logically deduce that 60% are as good or better!

EXCHANGE RATES

The difference in currency from one country to the next can cause many problems for consumers. There is also a variation from one day to the next. Some currency exchange companies may charge an extra amount; this is referred to as commission. Being aware of these facts allows the consumer to correctly evaluate the relative amount they are spending while abroad. They need to ask themselves, "Which is the more expensive: a coat costing $10 or one costing 15 euros?" The concept of ratios can be used to solve this particular problem: If that day's ratio is $1 to 1.2 euros, then $10 = 12 euros. Hence, the $10 coat is the better deal. Obviously, it pays to be aware of exchange rates.

PHONE COMPANIES

It can be difficult choosing the best company to use for a mobile phone. They all offer different rates and different incentives. A graph is a good way to compare different phone options. It may save money in the long run. For example, company A has a fixed charge of $20, and charges $1 for every 10 minutes; company B has no fixed charge, but charges $1 for every five minutes for the first two hours and then $3 every 5 minutes thereafter. Figure 1 shows a comparison graph. If the consumer uses the phone for less than 130 minutes a month, then option A is the better deal; otherwise company B offers the better deal.

TRAVEL AND RACING

Before setting out on a trip, it is important to assess travel times. To work out how long a 100 kilometer journey would take, one could make an approximation of 80 km/hour, which would therefore make the trip take 1 1/4 hours.

Another example is a man taking part in a rally. The overall length is 120 kilometers. He completes the first 60 kilometers in 1 hour and twelve minutes. To win the prize he needs to average over 100 kilometers an hour for the whole race. It would be impossible, because even if he travels at phenomenal speeds, he still wouldn't get his average speed above 100 kilometers an hour. In fact, even assuming he could arrive at the finishing post instantaneously, he still would only match the target, not beat it.

PROPORTION AND INVERSE PROPORTION

Many problems in real life have simple proportional laws and so are easy to solve. If 10 people on average can produce a factory output of 1,000 units, then 20 people on average should be able to produce 2,000 units. This deduction is called direct proportion. Unfortunately, it is not always that simple; careful reasoning is required before stating what could be the wrong solution.Suppose it takes 10 people 10 hours to do a job. How long would it take two people? The answer is not two hours! There are less people and so the job should take longer. This

particular case is an example of inverse proportion. It can be worked out using the unitary method: 10 people: 10 hours; 1 person: 100 hours; 2 people: 50 hours.

Even though proportion appears easy, when it is applied to other real-life problems it can get much more complex. For example, a company is producing boxes for storing model cars. The boxes are 2 cm by 2 cm by 2 cm. For a special edition, they want to create a box with a volume that is twice as big. What should the length of the sides be? The apparently obvious, yet incorrect, answer is for the sides to be 4 cm long. But the 2 cm sides give a volume of 8 cm^3, while the 4 cm sides give a volume of 64 cm^3. Much too big! By doubling the sides, the volume becomes 8 times as big. This is called cubic proportion.

If solving a problem that involves proportion, it should be determined whether it is direct proportion or not. It is also a good idea to always check answers afterwards.

ECOLOGY

A problem facing ecologists at the moment is the saving of endangered animals. Statements are frequently made concerning those dwindling in stock, and radical solutions are suggested. Yet, it is essential that the solutions be explored before any action is taken.

To model situations encountered in ecology, mathematical equations are set up that are indicative of the way the population changes as time progresses. These can be referred to as differential equations. These indicate how a population continually changes from second to second. This can be a bad model for species that breed at specific times. Such a population will have very distinct, regular changes.

The type of equation used to solve these situations can be known as difference equations. This would be used to illustrate changes over discrete periods of time. A list of equations, often referred to as a series of equations, is produced. These equations would each correspond to a different variable within the ecosystem in question. These are then solved, often using computers, to suggest the outcomes if different methods are used. If an equation is solved using computers, it is often referred to as an analytical solution.

A simple example to consider is that of rabbits and foxes. The ecologist will consider that the more rabbits there are, the quicker they will breed and hence the population will increase. If there are more rabbits, there is more for the foxes to eat, and so the foxes thrive and their population increases. Conversely, more rabbits are eaten, so their population decreases. Each of these lines could be represented by an equation and these could be used as indications of how the populations will develop.

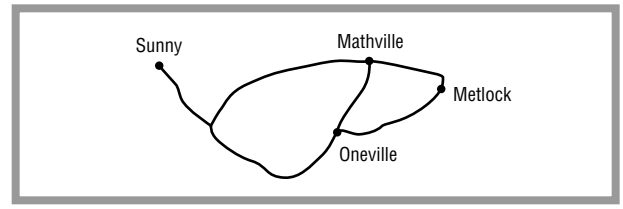


Figure 2.

TRANSLATION

As the commercial possibilities expand, and more and more cultures mix and work together, the ability to communicate is becoming increasing essential. Yet it is virtually impossible for a human translator to be present at all times to assist between different languages. It is for this reason, as well as cost consideration, that the concept of computerized translation is very appealing. Yet the ability to turn a random phrase in English into Spanish is difficult, if it is to be done efficiently. The simplest solution would be to have all conceivable phrases stored somewhere for each language, and to then link them. This is often called a one-to-one (functional) solution.

Careful consideration should, however, reveal the limitations of such an idea. The number of possible sentences in a language is unimaginably vast. The aim is therefore to program the computer with a sense of grammar and language structure. When a sentence is typed in, the computer recognizes whether words are verbs, nouns, or prepositions, converts these into the required counterpart, and then applies the correct grammar. This in itself is a remarkably complex task. Computers are still poor translators. However, the continual development of computers is allowing advances in such areas.

NAVIGATION

Strictly speaking, for many transportation companies, navigation is concerned with getting from point A to point B in the shortest time and cheapest way possible. A company will set out with the sole objective of finding this route. Finding the shortest distance is a large discipline of mathematics and often goes under the overall umbrella of decision mathematics.

To solve this problem, the company would make a map indicating all the possible routes and their respective costs. Figure 2 is an example of such a simplification.

Paradox

A paradox is a statement that seems to contradict expected reasoning. There are many famous paradoxes within mathematics and they often lead to exploration into new areas to try to evaluate why they occur. For example, the Sorites paradox. Sorites is Greek for heap and describes a set of thinking problems. At what point does a pile of sand denote a mound of sand? One grain clearly isn't a mound; add one more grain to this, and little difference has been made. By this definition, adding one grain each time still means there is no mound. At what point is a mound achieved? Conversely, if there is a mound and a grain of sand is removed, there is still presumably a mound. Keep removing one grain, and when is there no mound? Is it just the limitations of language that cause the apparent paradox?

Another paradox, originally expressed in ancient Greek, is well-known. A man fires an arrow at a moving target, albeit one that is slower than the arrow. Unfortunately, the arrow never hits the target. This is because by the time the arrow would have caught up with the target, this object has moved that much further on. So the arrow needs to travel a bit further, but by this time the target has once again moved. And so the argument persists. This entire argument has now been resolved and indeed is linked to a whole area of mathematics often referred to as convergence and divergence in sequences. These are extremely important areas in number theory.

Cost is a generic term used to denote the area of consideration. This could be time, or distance, or cost, or even gasoline consumption. An algorithm is then used to solve this problem. There are many methods available; the main one used is called Dijkstra's algorithm. Any electronic route-finder on cars will probably apply this method. A more complete algorithm used is called Floyd's. This is a repeated version of Dijkstra and finds the shortest distance between all points on a map.

The maps used are always simplified versions of the real-life situation. They will never resemble visually the actual physical situation. These maps are referred to as graphs, the roads are often called arcs, and the places where roads diverge or converge are called nodes, or vertices. This leads to a large area of real-life mathematics called graph theory.

GRAPH THEORY

Graph theory is often used to solve real-life problems, often those expressed in words that appear complex on the face of it. For example, the problem is to find the most efficient way to build a car using the minimal number of people, while completing the task within a prescribed time. The way to solve this problem is to identify the tasks and to construct a precedence diagram for the situation. The diagram merely indicates the order in which certain tasks need to be performed. It is obvious, for instance, that the engine cannot be placed in the car before the car itself has been built.

The situation thus described would then be solved using a method often called critical path analysis. Diagrams to show number of workers can also be drawn, which show how many people are required at any one time and would be used during the hiring process and to plan wages. These concepts are important to learn when considering a career in management and business.

LINEAR PROGRAMMING

Linear programming is used to solve such problems as how to maximize profit and minimize costs. The situation is simplified into a series of simple equations, and these are solved to present the optimal, or best, solution. For example, a company wants to produce two items of candy. Candy A will sell for \$1.50; Candy B will sell for \$2. The company wants to produce at most 1,000,000 candy bars altogether. Due to demand, it wants to make at least twice as much of A as of B. The ratio of the secret ingredient X in the two candy bars is 2:5. The company has 7,000,000 parts of ingredient X. How much of each should they produce to maximize profit?

The problem is solved as follows: They let x = the amount of Candy A made, and y = amount of Candy B made. Then, they want to maximize $1.50x + 2y$, since this denotes profit, subject to: $x + y < 1{,}000{,}000$ (total number of bars less than one million); $x > 2y$ i.e. $x - 2y > 0$ (twice as many of A as of B); $2x + 5y < 7{,}000{,}000$ (Total amount of ingredient X is less than 7,000,000 parts).

These equations can then be solved to find the optimal solution. They can be expressed graphically, using x- and y-coordinates to represent amount of candy A and candy B. These equations are linear because the coefficient of both x and y is 1. It is best solved using a computer. A method that is most efficient is called the simplex method. A computer is able to use the algorithm quickly and give the optimal solution in virtually no time at all.

TRAVELING SALESPERSON

Most companies need to travel either to market their product or to make deliveries. It is essential that this be done as efficiently as possible. Often a delivery will do a circular trip, calling at all required places. To save gas, the shortest route is found, though this may be in terms of time, or gas, or cost, or a combination of many factors. This requires graph theory to find a solution. Nodes are drawn to represent the places required and arcs are used to represent possible journeys.

There is no easy way to find an optimal solution. For extremely large routes, even a computer would take years to reach an optimal solution. For this reason, a trial and improvement technique is used. This is an important concept in mathematics. Estimates for worst-case and best-case scenarios are found. A logical search (often referred to as an inductive process) must take place. Gradually, improvements are made, until the company is satisfied with the solution. They may stumble upon a better solution later. The company that achieves the better solution will be the one that survives.

POSTMAN

A mailman who needs to walk down all streets in a particular precinct will want to take the shortest route possible, and avoid repetitions, if possible. Consider Figures 3 and 4.

In Figure 3, all of the roads (arcs) are complete. However, Figure 4 has one of the roads (arcs) removed. Even though there are fewer "roads" to go down, the actual solution takes longer to perform. It is actually the case that a good solution exists if all nodes have an even number of roads/arcs leading out of them. If there is a node with an odd number of roads coming out of them, then the problem becomes more complex.

To solve the problem, a consideration is taken of the odd nodes. As a reminder, this means the nodes with an odd number of roads coming out of them. The shortest arcs between such nodes are then doubled up. This is equivalent to walking up and down the road twice. It is like meeting a dead-end and the postman has to double back.

There are many different jobs where such analysis is required. Many bulk delivery firms will use such ideas. It can also be used for hypothetical problems such as where the arcs represent tasks and where all the tasks need to be performed, though not in any particular order.

ROTA AND TIMETABLES

One of the more complex aspects of any business is that of staffing levels and evaluating when staff should work. Many food outlets require shift patterns to be established, and the average high school will have many hundreds of teachers that need to be organized. A careful, logical approach is required to meet the demands.

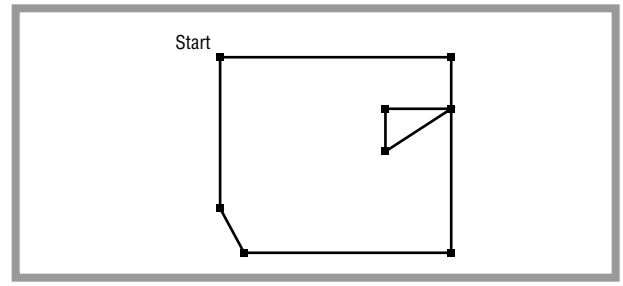


Figure 3.

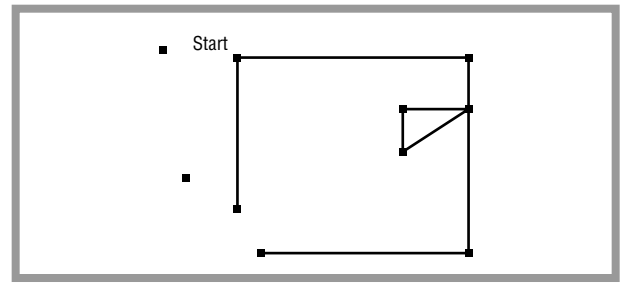


Figure 4.

SHORTEST LINKS TO ESTABLISH ELECTRICITY TO A WHOLE TOWN

What is the most efficient way to connect a whole town to a main electricity supply? Clearly, the most efficient solution would be the one using the smallest length of cable. There are two established techniques for solving this problem.

Drawing a graph is required to solve this problem. Nodes are used to represent houses, and arcs are used to represent all the possible connections available. The graph will be a complete graph. This is because all the different possibilities will be considered. One of the two following efficient methods will solve the problem.

In the Kruskal's algorithm, all the different possible cable lines are ranked from shortest (best) to longest (worst). Then cable is progressively added in until all the houses are connected. In the Prim's algorithm, it is the houses that are progressively joined by lengths of cable. Starting with the house that is closest each time, all houses are joined together.

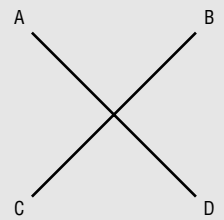
RANKING TEST SCORES

Ranking a long list of numbers occurs often in real life. This seems like a trivial task until there is a list of

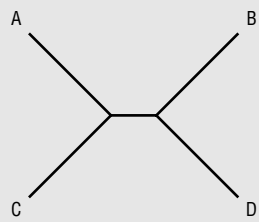
Connecting Four Towns

Consider four towns, each located at a vertex. A rail network is required to connect these four towns. Which of the following two solutions, Option A or Option B, is the optimal solution?

It turns out that Option B is the better solution. Indeed, by formulating mathematical expressions for the railway tracks, calculus can be used to evaluate what the length of the horizontal section must be for the smallest route. This will depend on the exact distances between each of the towns.



Option A.



Option B.

substantial size. Suddenly, a logical method is required. There are many different methods used, all going under the name of sorting algorithms. They all have different advantages and disadvantages. These algorithms may be programmed into software to allow computers to do the hard work. A computer needs an explicit set of instructions if it is to complete a task. The programmer must consider the amount of coding required to get the sort function to work.

SEARCHING IN AN INDEX

With a lot of information, it can be difficult to find one precise piece. It is for this reason that a dictionary is ordered sequentially. In another example, a student may have a large amount of school notes, each page numbered and in order, and the student needs to find a specific page to study for an exam. The method to use is called binary search.

This method requires a numbered list. This would be the case in most examples of filing. A good starting point would be the halfway point in the list. The student can look through the upper half first, then the lower half, until the specific page is found. This is much quicker method than randomly looking at pages. Obviously, a computer would be much quicker!

EFFICIENT PACKING AND ORGANIZATION

To pack the most objects in a given space requires careful mathematics. One method is extremely good at these packing situations. The rule is to order the objects first, from largest to smallest, and then pack them in that particular order.

SEEDING IN TOURNAMENTS

One of the prerequisites for many sporting events is that the best players don't meet each other until the later stages of the game. To accomplish this, players are allocated seeds, or rankings based upon their past and current performance. The players are then often pooled into different groups and the fixtures are arranged initially within groups. This will ensure that seed 1 and seed 2 will not meet until later in the tournament.

ARCHITECTURE

Buildings must be designed by taking several factors into consideration. It is to resolve the myriad issues that architectural design is so important. Architects are workers with a fully functional knowledge of the mathematics behind construction.

Objects of such magnitude as buildings must be constructed of materials that support the extreme forces exerted on them. The tensile strength of a material involves how much it can be stretched without deforming. The compressive strength corresponds to the ability to withstand compressive forces. (It would be disastrous if the walls of a building began to shrink!)

The shape and structure of the building is also important. Certain configurations are recognized as having a much greater stability. Often, geometry will be used to ensure that angles of adjoining structures maximize the strength required.

COOKING INSTRUCTIONS

Many meals require precise instructions, depending on oven type and power. It is then up to the consumer to evaluate the cooking time for the product. Many pieces of meat have times prescribed according to mass. For example, a chicken may require 30 minutes cooking, plus an extra 30 minutes per 500 grams. It is obviously important to be able to understand such instructions.

RECIPES

Recipes are real-life examples of word problems. They provide exact quantities to make a meal for a specific number of people. It is then up to the individual to adjust the ratio accordingly. This is an example of direct proportion. It is an essential skill for those involved in mass catering or indeed in any production to be able to scale up required ingredients to satisfy variable orders.

LOTTERIES AND GAMBLING

Many millions of people gamble every day. They are often enticed by vocabulary, such as even chance or good chance, without really knowing what the phrases mean. The odds in horse racing always start as a ratio; it is up to the betters to understand the relative merits of the odds and make a judgment accordingly.

BANKS, INTEREST RATES, AND INTRODUCTORY RATES

The modern banking market is extremely competitive. One of the main concerns when establishing a savings account is that of interest. Each bank may offer a slightly different level, and some offer initial rates that soon change.

There are two different types of interest. The main type is called compound interest. This is normally paid yearly and is evaluated from the amount currently in the account. The second type is simple interest. This is a fixed amount. It is often worked out by looking at the initial amount deposited into the account.

An example would be look at savings account A, which has an initial deposit of $1,000 that offered a yearly interest rate of $100 fixed; savings account B offered 8% yearly. The progression of account A would be 1,000, 1,100, 1,200, 1,300, 1,400, 1,500, 1,600, 1,700; the progression of account B would be 1,000, 1,080, 1,166, 1,259, 1,360, 1,469, 1,586, 1,713. Clearly, option B is relatively slower to start off with. However, after seven years the amount in account B overtakes that in account A. It is always important to look in detail at a mathematical situation and not just take a short-sighted view of the problem.

FINANCE

A company will often lay down objectives for the forthcoming year. These will be in the form of a business plan that describes the growth desired and what expenditures can be used, among other factors. It is often up to consultants to suggest ideas for how such objectives can be achieved. Economics can be modeled through a range of equations and economic principles are often applied to the stock market and growth of countries and cities. A consultant would be able to use the initial data and work out the best way the resources can be used to ensure the company achieves good results.

The study of economics is highly mathematical. There are many accepted models used within the business world.

DISEASE CONTROL

Many scientists currently monitor disease and try to evaluate likely outbreaks. The World Health Organization (WHO) may be interested in the likelihood of an outbreak of malaria in a certain part of Africa. Mathematical models are constructed, using data available, to evaluate possibilities. These models will frequently involve past data, as well as expected data. Understanding the probabilities of recurrences and the likelihood of location would be a useful tool in combating the many serious diseases.

GEOLOGY

Geology is the study of the physical Earth, and most aspects would be considered relevant to the real world. As of 2005, due to the Asian tsunami disaster occurring in 2004, an awareness of the forces of nature is at the forefront of people's consciousness. The question that many officials may ask is "Will this happen again?" or "When would such an occurrence happen?" or "How would a tsunami affect us if it occurred closer to our country?"

The mathematician would work out the many different possibilities that could occur. Perhaps by studying the effects of the recent disaster more information will be accessible and further developments made. Yet to do this, it would be broken down into the following key areas, such as where could such an event occur, how unstable is the area, how deep are the oceans, and what effect would this have?

The mathematician would then be able to apply models to each of these situations and produce a logical answer giving the range of expected possibilities. The study of dynamics, especially in fluids such as the oceans, is a vast area of applied mathematics. Many famous mathematicians (for example, Euler) spend years of their life studying such issues.

SURVEYING

When building on a new site, a company would first of all be expected to analyze the area to ensure no dangers are around. Yet to solve this, consideration would have to be taken into what safe actually means within the context, and compare it to the construction being built. The situation would be simplified into key areas, including what sort of weight can the land tolerate and what effect on the environment would the project have? Such questions would be explored mathematically through a consideration of the weight of the engineering project and the stability of the surface.

STORE ASSISTANTS

Store assistants are constantly faced with word problems that may need immediate response. A customer may ask how much a group of items would cost and the assistant may not have a calculator at hand. The sales assistant must be able to give an immediate response.

STOCK KEEPING

Store managers must work out how much stock to order. If too much is ordered, it may be wasted; yet if too little is ordered, customers will be dissatisfied. Managers develop their own techniques for solving such questions, however much of what they do will depend upon instinct and experience. Many real word problems require experience to be solved. This can be paralleled in pure mathematics. A good store manager will analyze sales of the same period for previous years. They will evaluate averages and use these figures to determine the amount that will be required. They may also produce graphs to show how the average amount is changing. These are referred to as moving average problems. For examples, average sales may have gone up by \$10, then \$20, then \$30; consequently, a fair estimate may be made that the next increase will be \$40. The manager then uses this figure when deciding how many units to order. Once again, the problem is solved through converting the real-life situation into exact mathematical figures. These allow for simple conclusions that can be backed up with fact.

ACCOUNTS AND VAT

Deciphering monetary information often requires a mathematical answer.VAT is a tax paid on items that are not essential and is required by law within the European Union. Any U.S. company selling into the EU has to, by law, charge VAT at the required level.

If an item's basic cost is known, then VAT is easy to work out. The tax is the required percentage of the total cost. For example, a coat exported to the United Kingdom cost \$85.11 before VAT was added. If the U.K. VAT is 17.5% then the cost of the coat (rounded in dollars) becomes $\$85.11 + \$85.11 \times (17.5/100) = \$100.00$. The person is able to claim the VAT tax of \$14.89 back from the U.K. government if the coat is essential for his employment.

BEARINGS AND DIRECTIONS OF TRAVEL

The shortest route between two points on a flat surface is the straight line connecting the two points. However, how is motion achieved in that straight line? This is a question that transport companies, especially nautical-related transport, need to consider all the time because other factors are continuously trying to influence the motion of the vessel. There will be currents and wind trying to steer the vessel off course. The ship would therefore have to steer a course that compensates for these extra factors. These problems can be solved using bearings and trigonometry. Today, of course, sensors will detect the forces present and computers will be able to adjust the steering as required.

QUALITY CONTROL

It is important for companies to monitor output to ensure that goods meet standards. The authorities often define these standards, and not meeting them could lead to heavy fines and/or closure. For example, the criteria are that only 5% of products are below a required size and the company produces one million of these items a day. How do they monitor their output?

A system is often used called systematic sampling. Every one hundredth item produced is checked against the required criteria. The company will then keep a running total of items failing or passing the test. As long as a

sufficient number is above the required standard, the company will keep producing. The authorities will normally publish guidelines, and the company uses those.

Sampling is used to solve a wide range of such problems. In different situations, different sampling techniques are used. Samples are used because it is often impossible to test or analyze every single item in a population.

WHAT IS THE AVERAGE HEIGHT IN A NEIGHBORHOOD?

Manufacturers of items ask this sort of question all the time when the size of people, for example, has direct relevance on production. It would be a bad business decision to produce small clothes if the population happened to be a tall one. Yet, how would a company evaluate the average height?

The company would first identify the target market. This is important if their line of production happens to be jackets for women. They would then need to pick a random sample, which reduces the potential for bias. Often the company will do a form of quota sampling. This is a method to ensure that people of all ages are picked. A quota is a group. The company will identify all the relevant groups and pick out a random people from each. The formula used to find the number of people in a random sample or quota group is normally the square root of the entire targeted population.

OPINION POLLS

Opinion polls are used to answer such questions as "Who is the most popular politician?" Politicians can use them as propaganda, in both a positive and negative way. Opinion polls, however, are often biased. Mathematically speaking, opinion polls are not necessarily considered to be sound. They frequently target only a select group in a population and thus lead to often conflicting and contradictory evidence.

WEATHER

Forecasts are used and needed across many spheres in many different occupations. It is not possible to say what will happen; instead forecasters deal with what is most likely to happen. The reason weather cannot be predicted with much accuracy is due to a mathematical idea called chaos theory. Basically, there are so many interactions happening at both the macroscopic and microscopic level that any slight perturbation in any of these interactions could seriously affect the weather's outcome. Many sporting events and agricultural areas rely exclusively on forecasts to plan their daily tasks.

Riddles

A riddle is a written or verbal statement that requires exact logic to solve. The answer should be unique and make exact sense; otherwise, it is insolvable. Riddles parallel a lot of work done in mathematics in real life. They require sentences to be simplified into understandable ideas. Solutions can then be posed, until the correct solution is acquired. The solution of a riddle mimics the modeling method in mathematics.

To solve a riddle, one must consider the set of solutions that solve each sentence. The solution that overlaps all parts of the riddle is the final solution. Consider the following challenging riddle:It is better than God and more evil than the devil.Rich people want it, poor people need it.You die if you eat it.

What is the riddle's solution? (The answer is "nothing.")

The fundamental concepts behind weather forecasting are the understanding of the interactions in the atmosphere and the modeling of this using mathematics. Powerful computers are today used to predict the likely outcome, churning out vast output of data. The art of predicting weather is often referred to as meteorology. It is certainly not an exact science. To try to get a realistic answer to the problem of weather forecasting, the super computers produce different outputs with a slightly different starting point (a forced perturbation). The average can then be taken. These small perturbations often lead to dramatic changes in the output. There is frequently a dramatic divergence in solutions, especially when one begins to predict more than just three or four days in the future.

THROWING A BALL

How one throws a ball to maximize the distance achieved is of particular relevance within the sporting world. The answer is solved through a series of assumptions. If it is assumed that the ball is thrown approximately from ground level and that the only force acting on the ball is gravity, the solution is that the angle should be 45°. It is clear why the angle affects the solution. If the ball is thrown vertically upwards, it will cover no distance, but if it is thrown horizontally, it will fall quickly to the ground. This model can then be improved and different solutions will be thus arrived. However, this gives the mathematician a starting point from which to develop a theory.

MEASURING THE HEIGHT OF WELL

The problem when constructing a working well for a village in Africa is that there is a chasm already present. There is a simple way to approximate its depth. If a stone is dropped down the well, the time taken to reach the bottom can be measured. A distinct sound would be heard as it hits the water. The depth of the well can be approximated using the formula: $d = 4.9 \times t^2$.

DECORATING

When setting out on a renovation project, one of the first questions will be a consideration of the materials required. To minimize the cost of decoration it would be advisable to use careful mathematics to evaluate the quantity of material required. A professional decorator will not want to mix a required hue only to find that there is not enough to finish the whole room.

These types of problems can be easily solved through a consideration of area. Rooms are generally regular. A simple calculation involving width and height would give the amount of wall space involved. The materials should have indications on the labels informing the consumer how much area they will cover. It is then a simple case of using proportion to evaluate the amount of material needed.

DOES GLOBAL WARMING EXIST?

There are many different arguments on either side of the debate of global warming. Mathematics provides a way of looking at such issues and problems in a non-emotive way, allowing for careful and logical reasoning. It is, however, easy to manipulate many ideas involved and the issue must be studied free from influence either political or otherwise. This underpins the mathematics behind independent surveys. It is a tool. Like all tools it can be used flexibly in ways that are not obvious to the layman.

DOES MMR (MEASLES, MUMPS, RUBELLA) IMMUNIZATION CAUSE AUTISM?

There is a reported link between immunization and subsequent disease. Mathematics, especially statistical ideology, is used to test the likelihood of such a link existing. Unfortunately, the mathematics is often lost beneath emotion and ideology until the evidence itself is discounted or stated to be invalid. This is the main reason why statistical tests used to investigate links need to be done as rigorously as possible. There will always be an element of doubt in the conclusions reached. The reduction of this doubt will lead to more convincing arguments, and so results can be displayed and credible conclusions reached. Recent research does not establish a link between MMR immunization and autism.

Potential Applications

The existence of word problems and their necessity within society will never cease. Language will continue to develop and so will the mathematical thirst to solve and to explain. The ability to solve such problems and the skills to explain in simple terms will always be considered an essential skill in all areas of employment.

As time passes, mathematical models will become more and more sophisticated and the advent of more powerful computing will allow more accurate solutions. More and more advanced questions about the universe and the inherent mathematics that underpins it will continue to be pursued. Who knows how far the solutions will take us?

Where to Learn More

Books

Parramore, K., J. Stephens, G. Rigby, and C. Compton. *MEI Decision and Discrete Mathematics* London: Hodder Arnold H&S, 2004.

Porkess, R., et al. *MEI Statistics 2* London: Hodder Arnold H&S, 2005.

Web sites

Value Added Tax. Online Resources. <http://www.vat.com/faq.html> (March 1, 2005).

Zero-Sum Games

Overview

A zero-sum game is a game in which whatever is lost by one player is gained by the other player or players. The study of zero-sum games is the foundation of game theory, which is a branch of mathematics devoted to decision-making in games.

In mathematics, all situations in which there are two or more parties—people, companies, teams, or nations—making decisions that affect some measurable outcome are "games." The decisions made by a game player make up that player's "strategy." The goal of game theory is to calculate the best strategy for a given game. Zero-sum games are a special part of game theory that can be applied in law, military strategy, biology, and economics.

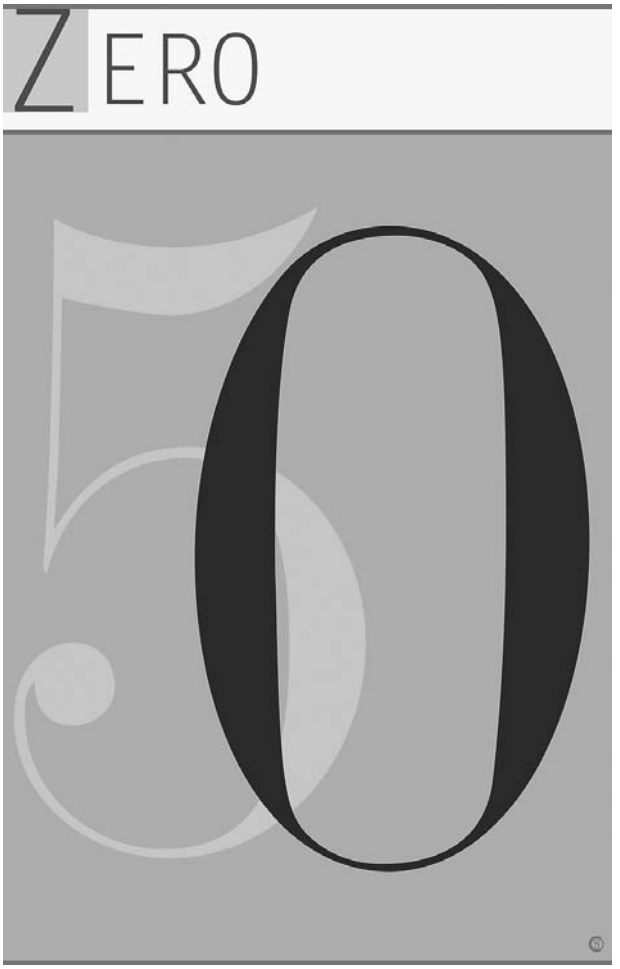
Games are not necessarily played for fun. They can be deadly serious. Chess, cards, and football are considered "games" in game theory, but so are business and war. Not all the pastimes we call "games" are games in the game-theory sense. The children's card game called War is an example of a game that is not a game (mathematically speaking). In War, the players repeatedly match cards, one from each player, and the player with the higher card takes the pair. They continue until one player holds all the cards. Which player ends up with all the cards depends only on how the cards have been shuffled and dealt. No decisions are made by either player, so there is no way to choose a strategy. The winner is decided by pure chance.

True games can, however, involve an element of chance. In football, for instance, a player can slip on wet turf, make a freak catch, or get confused and throw the ball the wrong way. Sometimes the winning team is even decided by such an event. But football coaches still plan strategies, and strategy does make a difference.

Fundamental Mathematical Concepts and Terms

In a zero-sum game, the players compete for shares of something that is in limited supply. One player's loss is the other player's gain: if your slice of pie is bigger, mine must be smaller.

The term "zero-sum" refers to the numbers that are assigned to different game endings. If winning a game of chess is assigned a value of $+1$, then losing a game has the value -1 and the sum of the loser's score and the winner's score for every game is $1 - 1 = 0$, "zero sum." When there is a draw, both players get 0 points and the game remains zero-sum because $0 + 0 = 0$.



In zero sum games, winners entail losers. STEVE COLLIER; COLLIER STUDIO/CORBIS.

Two-player zero-sum games are also called strictly competitive games. Games may also have more than two players, as in poker or Monopoly. When three or more players play a zero-sum game, some players may team up or collaborate against the others, so multi-player zero-sum games are not "strictly competitive."

The theory of zero-sum games is the starting point for the theory of all other games, which can be lumped under the term "non-zero-sum games." Non-zero-sum games are games which are not played for fixed stakes.

The most famous non-zero-sum game is the Prisoner's Dilemma, first proposed by Merrill Flood and Melvin Dresher at the Rand Corporation in 1950. In this situation, there are two prisoners who have committed a serious crime. The police put each prisoner in a separate cell and try to get them to confess by telling each prisoner (falsely) that the other prisoner has already confessed, and that if they will also confess, they will get a reduced sentence. But, the police add, if the prisoner does not confess, they will get a heavy sentence.

If both prisoners confess, they will both get reduced sentences. If only one confesses, then the one that confesses will get a reduced sentence and the other will get a heavy sentence. If neither confesses, then both will be freed. Obviously, it would be best for both prisoners if they refused to confess. Yet, it can be shown by game theory that the most mathematically "rational" thing for each prisoner to do is to confess. This is a "dilemma" or no-win situation because the best strategy is to confess and take a reduced sentence rather than to refuse to confess, because each prisoner cannot guarantee what the other will do. Though not confessing might result in no sentence at all, a heavy sentence could result for a prisoner who does not confess when the other does. The guessing game played by the two prisoners is a non-zero-sum game because both prisoners might win (go free) or lose (get sentences) at the same time: there is not a fixed number of years of imprisonment that must be divided between the prisoners.

Real-life Applications

GAMBLING

Competitive gambling for money is usually a zero-sum game because the money won by one player must be lost by another. There is a fixed amount of money, and rolling dice or dealing cards cannot destroy it or create any more. Zero-sum game theory can therefore be used to find the best possible strategies for such games. This applies to games in which there is an element of choice or strategy, such as poker. In fact, the game of poker was what inspired Hungarian-born American mathematician John Von Neumann (1903–1957) to invent modern game theory, which he did starting with his 1928 article, "Theory of Parlor Games."

However, not all gambling games are "games" in the game-theory sense. Playing a slot machine is not a game, for example, because it is a matter of pure chance, all the player does is pull the handle or push the button. Game theory has nothing to say about activities like slots, roulette, dice, or lotteries because they allow no choices to the player and therefore no strategy. The only choice the player has is to play or not play. Mathematics can deal with games of pure chance, but this is done using probability theory, not game theory. Probability theory is used in game theory to deal with games that mix strategy with chance.

EXPERIMENTAL GAMING

Psychologists have used game theory to study how human beings make real-world decisions. They do this by asking volunteers to play a game. The psychologists use game theory to calculate the best or optimal strategy for the game and compare the behavior of the volunteers to the results of game theory. Psychologists have studied behavior in both zero-sum and non-zero-sum situations. They have often found that people do not behave in the way that game theory says is most "rational."

This does not necessarily mean, however, that people act foolishly. People may simply disagree with the mathematical definition of rationality. For example, if people are offered an (imaginary) choice of $1,000 in cash or a black box that has a 50% chance of containing either nothing or $10,000, they usually take the cash. Mathematics, however, says that the player's most "rational" choice is to maximize their expected or average winnings by choosing the black box. If the game were played many times over, a player who always chose the box would make more money on average (about $5 thousand) than a player who always took the $1 thousand. In this sense it is more "rational" to take the box.

But there is something artificial about saying that the behavior of a player who takes the cash is not rational. Why should a person take a 50% chance of getting nothing when they could get money without risk? This desire to avoid drastic risk is an example of what game theorists call "risk aversion." People usually prefer a strategy that protects them from disaster to a strategy that offers them big potential winnings but exposes them to possible disaster.

CURRENCY, FUTURES, AND STOCK MARKETS

Currency and futures trading are zero-sum games. Currency trading is a form of money investment in which speculators buy up one kind of money—dollars, pounds, euros, yen, or other—and then sell it again, trying to make a profit. For example, if 1 US dollar can buy 1.01 euros in Germany, and 1.01 euros can buy 1.02 yen in Japan, and 1.02 yen can buy 1.03 dollars in the U.S., then an investor can make $.03 by taking $1, buying a euro with it, buying a yen with the euro, and buying a dollar with the yen. This would be a way of getting something for nothing, except that for every penny made in the currency-trading market somebody loses a penny in the currency-trading market. The market does not generate new wealth: like a poker game, it only moves money around. Currency trading is therefore a zero-sum game. In addition, such trading as outlined above does not take into account fees that brokers charge to make transactions.

In futures trading, speculators gamble on whether unprocessed commodities like grain, beef, or oil will be worth more, less, or the same in the near future. Since a loss for the seller of the commodity is a gain for the buyer of the commodity and vice versa, the futures market is also a form of a zero-sum game. The commodity markets allow producers to fix sale prices ahead of delivery and therefore manage their risk of losing money.

There is debate about whether the stock market is a zero-sum game, but most economists agree that it is not. In the stock market, investors buy shares of ownership in companies. For instance, buying a single share might make you the owner of one millionth of the ABC Corporation. These shares can be bought and sold. As long as the value of the companies being owned remains fixed, buying and selling stock in them is a zero-sum game; however, the companies are real-world enterprises that may decrease or increase in value. Demand for a product might increase or decrease, or a vital resource (like oil) might run out, a company might go out of business, or a new technology might be developed that increases productivity and makes more real wealth. Any of these events changes the amount of wealth that the stock-market game is being played for.

WAR

War as such is not a zero-sum game. In almost any case, if both sides helped each other instead of fighting, they would be better off than if they fought. And, if the war is destructive enough, both sides, even the "winner," may end up worse off than before.

However, particular battles are often zero-sum games. The military forces fighting a battle are trying to destroy each other's resources—to kill soldiers and to destroy weapons, vehicles, and supplies. A loss for one side is a gain for the other, which is the primary feature of zero-sum games. Military strategists do in fact study battle strategy in terms of zero-sum games as well as in terms of more complex, non-zero-sum game theory.

Where to Learn More

Books

Colman, Andrew M. *Game Theory and Its Applications in the Social and Biological Sciences.* New York: Routledge, 1999.

Davis, Morton D. *Game Theory: A Nontechnical Introduction.* New York: Basic Books, 1970.

Straffin, Philip D. *Game Theory and Strategy.* Washington, DC: Mathematical Association of America, 1993.

Glossary

80/20 rule: A general statement summing up the tendency for a few items to consume a disproportionate share of resources, such as cases in which 20% of a store's customers lodge 80% of the total complaints.

Acceleration: A change of velocity (either in magnitude or direction).

Actuary: A mathematical expert who evaluates the statistical likelihood of various insurable events for underwriting purposes.

Algebra: A collection of rules: rules for translating words into the symbolic notation of mathematics, rules for formulating mathematical statements using symbolic notation, and rules for rewriting mathematical statements in a manner that leaves their truth unchanged.

Algorithm: A set of mathematical steps used as a group to solve a problem.

Analogue: A continuously variable medium, for use as a method of storing, processing, or transmitting information.

Analytic geometry: A branch of mathematics that uses algebraic equations to describe the size and position of geometric figures on a coordinate system. Developed during the seventeenth century, it is also known as Cartesian geometry or coordinate geometry. The use of a coordinate system to relate geometric points to real numbers is the central idea of analytic geometry. By defining each point with a unique set of real numbers, geometric figures such as lines, circles, and conics can be described with algebraic equations. Analytic geometry has found important applications in science and industry alike.

Angle: A geometric figure formed by two lines diverging from a common point or two planes diverging from a common line often measured in degrees.

Area: The measurement of a surface bounded by a set of curves as measured in square units.

Arithmetic: The study of the basic mathematical operations performed on numbers.

Array: A rectangular arrangement of numerical data in rows and columns, as in a matrix.

Average: A numeral that expresses a set of numbers as a single quantity. It is the sum of the numbers divided by the number of numbers in the set.

Axis: Lines labeled with numbers that are used to locate a coordinate.

Balance: An amount left over, such as the portion of a credit card bill that remains unpaid and is carried over until the following billing period.

Bankruptcy: A legal declaration that one's debts are larger than one's assets; in common language, when one is unable to pay his bills and seeks relief from the legal system.

Bicentric perspective: Perspective illustrated from two separate viewing points.

Binary code: A string of zeros and ones used to represent most information in computers.

Bit: The smallest unit of storage in computers. A bit stores binary values.

Boolean algebra: The algebra of logic. Named after English mathematician George Boole, who was the first to apply algebraic techniques to logical methodology. Boole showed that logical propositions and their connectives could be expressed in the language of set theory.

Bouncing a check: The result of writing a check without adequate funds in the checking account, in which the bank declines to pay the check. Fees and penalties are normally imposed on the check writer.

Byte: A byte is a group of eight bits.

Calculator: A tool for performing mathematical operations on numbers.

Calculus: A branch of mathematics that deals with the way that relationships between certain sets (or functions) are affected by tiny changes in one of their variables.

Cartesian coordinate: A coordinate system where the axes are at 90 degrees to each other, with the x axis along the horizontal.

Centric perspective: Perspective illustrated from a single viewing point.

Chi-square test: The most commonly used method for comparing frequencies or proportions. It is a statistical test used to determine if observed data deviate from those expected under a particular hypothesis. The chi-square test is also referred to as a test of a measure of fit or "goodness of fit" between data. Typically, the hypothesis tested is whether or not two samples are different enough in a particular characteristic to be considered members of different populations. Chi-square analysis belongs to the family of univariate analysis, i.e., those tests that evaluate the possible effect of one variable (often called the independent variable) upon an outcome (often called the dependent variable).

Chord: A straight line connecting any two points on a curve.

Coefficient: A coefficient is any part of a term, except the whole, where term means an adding of an algebraic expression (taking addition to include subtraction as is usually done in algebra). Most commonly, however, the word coefficient refers to what is, strictly speaking, the numerical coefficient. Thus, the numerical coefficients of the expression $5xy^2$ $3x + 2y$ are considered to be 5, -3, and $+2$. In many formulas, especially in statistics, certain numbers are considered coefficients, such as correlation coefficients.

Combinatorics: The study of combining objects by various rules to create new arrangements of objects. The objects can be anything from points and numbers to apples and oranges. Combinatorics, like algebra, numerical analysis and topology, is an important branch of mathematics. Examples of combinatorial questions are whether we can make a certain arrangement, how many arrangements can be made, and what is the best arrangement for a set of objects. Combinatorics can be grouped into two categories: enumeration, which is the study of counting and arranging objects; and graph theory, or the study of graphs. Combinatorics makes important contributions to fields such as computer science, operations research, probability theory, and cryptology.

Common denominator: A common denominator for a set of fractions is simply the same (common) lower symbol (denominator). In practice the common denominator is chosen to be a number that is divisible by all of the denominators in an addition or subtraction problem. Thus for the fractions 2/3, 1/10, and 7/15, a common denominator is 30. Other common denominators are 60, 90, etc. The smallest of the common denominators is 30 and so it is called the least common denominator.

Complex numbers: Complex numbers are so called because they are made up of two parts which cannot be combined. Even though the parts are joined by a plus sign, the addition cannot be performed. The expression must be left as an indicated sum.

Concentration: The ratio of one substance mixed into another substance.

Congruent: Two triangles are congruent if they are alike in every geometric respect except, perhaps, one. That one possible exception is in the triangle's "handedness." There are only six parts of a triangle that can be seen and measured: the three angles and the three sides. The six features of a triangle are all involved with congruence.

Conic section: The plane curve formed by the intersection of a plane and a right-circular, two-napped cone.

Constant: A value that does not change.

Convenience sampling: Sampling done based on the easy availability of the elements.

Coordinate: A set of two or more numbers or letters used to locate a point in space. For example, in two dimensions a coordinate is written as (x,y).

Cross-section: The two-dimensional figure outlined by slicing a three-dimensional object.

Cubed root: The relation of the volume of a cube to one of its edges.

Cubic equation: A cubic equation is one of the forms of $ax^3+bx^2+cx+d = 0$ where a,b,c, and d are real numbers.

Curve: A curved or straight geometric element generated by a moving point that has extension only along the one-dimensional path of the point.

Data point: A point in a graph or other display that depicts a specific value given by a function or calculation.

Decimal: Relating to the base power of ten.

Decimal fraction: A numeral that uses the numeration system, based on ten, to represent fractional numbers. For example, a decimal fraction for 2 and 1/4 is 2.25.

Decimal number system: A base-10 number system that requires ten different digits to represent numbers (0 through 9) where the value of a number is defined by its place (a place value system where a "1" could be valued at "one," "ten," "one hundred," "one thousand," etc.).

Decryption: The process of using a mathematical algorithm to return an encrypted message to its original form.

Degree: The word "degree" as used in algebra refers to a property of polynomials. The degree of a polynomial in one variable (a monomial), such as $5x^3$, is the exponent, 3, of the variable. The degree of a monomial involving more than one variable, such as $3x^2y$, is the sum of the exponents; in this case, $2 + 1 = 3$.

Dependent variable: What is being modeled; the output that results from a function or calculation.

Derivative: The limiting value of the ratio expressing a change in a particular function that corresponds to a change in its independent variable. Also, the instantaneous rate of change or the slope of the line tangent to a graph of a function at a given point.

Differentiate: The process of determining the derivative or differential of a particular function.

Digital: Of or relating to data in the form of numerical digits.

Dimension: The number of unique directions it is possible for a point to move in space. The world is normally thought of as having three dimensions. Flat surfaces have two

dimensional and more advanced physical concepts that require the use of more than three dimensions.

Distributive property: The distributive property states that the multiplication "distributes" over addition. Thus $a \times (b + c) = a \times b + a \times c$ and $(b + c) \times a = b \times a + c \times a$ for all real or complex numbers a, b, and c.

Dividend: A mathematical term for the beginning value in a division equation, literally the quantity to be divided. Also a financial term referring to company earnings which are to be distributed to, or divided among, the firm's owners.

Divisibility: The ability to divide a number by another number without leaving a remainder.

Domain: The domain of a relation is the set that contains all the first elements, x, from the ordered pairs (x,y) that make up the relation. In mathematics, a relation is defined as a set of ordered pairs (x,y) for which each y depends on x in a predetermined way. If x represents an element from the set X, and y represents an element from the set Y, the Cartesian product of X and Y is the set of all possible ordered pairs (x,y) that can be formed.

Encryption: Using a mathematical algorithm to code a message or make it unintelligible.

Enumeration: The study of counting and arranging objects.

Equation: A mathematical statement involving an equal sign.

Equivalent fractions: Two fractions are equivalent if they stand for the same number (that is, if they are equal). The fractions 1/2 and 2/4 are equivalent.

Estimation: A process that arrives at an answer that approximates the correct answer.

Exponent: Also referred to as a power, a symbol written above and to the right of a quantity to indicate how many times the quantity is multiplied by itself.

Exponential growth: A growth process in which a number grows proportional to its size. Examples include viruses, animal populations, and compound interest paid on bank deposits. The rate of growth is proportional to the size of the sample or population (i.e., a relation between the size of the dependent variable and rate of growth).

Fibonacci numbers: The numbers in the series, 1, 1, 2, 3, 5, 8, 13, 21, 34, 55, 89, 144 . . . , which are formed by adding the two previous numbers together.

Formula: A general fact, rule, or principle expressed using mathematical symbols.

Fractal: A self-similar shape that is repeated over and over to form a complex shape.

Fraction: The quotient of two quantities, such as 1/4.

Frequency: Number of times that a repeated event occurs in a given time period, typically within one second.

Function: A mathematical relationship between two sets of real numbers. These sets of numbers are related to each other by a rule that assigns each value from one set to exactly one value in the other set. The standard notation for a function $y = f(x)$, developed in the eighteenth century, is read "y equals f of x." Other representations of functions include graphs and tables. Functions are classified by the types of rules which govern their relationships.

Gambling: A popular form of entertainment in which players select one of several possible outcomes and wager money on that outcome.

Game theory: A branch of mathematics concerned with the analysis of conflict situations. It involves determining a strategy for a given situation and the costs or benefits realized by using the strategy. First developed in the early twentieth century, it was originally applied to parlor games such as bridge, chess, and poker. Now, game theory is applied to a wide range of subjects such as economics, behavioral sciences, sociology, military science, and political science.

Geometry: A fundamental branch of mathematics that deals with the measurement, properties, and relationships of points, lines, angles, surfaces, and solids.

Golden ratio: The number 1.61538 that is found in many places in nature.

Greatest common divisor: The largest number that is a divisor of two numbers.

Hypotenuse: The longest leg of a right triangle, located opposite the right angle.

Improper fraction: A fraction whose value is greater than or equal to 1.

Independent variable: Input data to a function. The input data used to develop a model where the outcomes or results are determined by function and/or calculation.

Inequality: A statement about the relative order of members of a set. For instance, if S is the set of positive integers, and the symbol $<$ is taken to mean less than, then the statement $5 < 6$ (read "5 is less than 6") is a true statement about the relative order of 5 and 6 within the set of positive integers.

Infinity The term infinity conveys the mathematical concept of large without bound, and is given the symbol ∞.

Inflation: A steady rise in prices, leading to reduced buying power for a given amount of currency.

Input: What is used to develop a model, the independent variables.

Integer: The positive and negative whole numbers. $-4, -3, -2, -1, 0, 1, 2, \ldots$ The name "integer" comes directly from the Latin word for "whole." The set of integers can be generated from the set of natural numbers by adding zero and the negatives of the natural numbers. To do this, one defines zero to be a number which, added to any number, equals the same number.

Integral: A quantity expressible in terms of integers (the positive and negative whole numbers). Also, a quantity representing a limiting process in which the domain of a function is divided into small units.

Integral calculus: A branch of mathematics used for purposes such as calculating such values as volumes displaced, distances traveled, or areas under a curve.

Interest: Money paid for a loan, or for the privilege of using another's money.

Irrational number: A number that cannot be expressed as a fraction, that is, it cannot be written as the quotient of two whole numbers. As a decimal, an irrational number is shown by an infinitely long non-repeating sequence of numbers. Examples of irrational numbers are pi (the ratio of circumference to diameter of a circle), e (base of the natural logarithms).

Iteration: Iteration consists of repeating an operation of a value obtained by the same operation. It is often used in making successive approximations, each one more accurate than the one that preceded it. One begins with an approximate solution and substitutes it into an appropriate formula to obtain a better approximation. This approximation is subsequently substituted into the same formula to arrive at a still better approximation, and so on.

Key: A number or set of numbers used for encryption or decryption of a message.

Knot theory: A branch of mathematics that studies the way that knots are formed.

Least-terms fraction: A fraction whose numerator and denominator do not have any factors in common. The fraction 2/3 is a least-terms fraction; the fraction 8/16 is not.

Line: A straight geometric element generated by a moving point that has extension only along the one-dimensional path of the point.

Linear algebra: Includes the topics of vector algebra, matrix algebra, and the theory of vector spaces. Linear algebra originated as the study of linear equations, including the solution of simultaneous linear equations. An equation is linear if no variable in it is multiplied by itself or any other variable. Thus, the equation $3x + 2y + z = 0$ is a linear equation in three variables.

Linear equation: An equation on which the left-hand side is made up of a sum of terms (each of which consists of a constant multiplying a variable), and the right-hand side which consists of a constant. For example, $2x_0 + 3x_1 = 4$.

Linear programming: A method of optimizing an outcome (e.g., profit) defined by a linear equation but constrained by a number of linear inequalities. The inequalities are recast as linear equation and the resulting system is solved using matrix algebra.

Logarithm: The power to which a base number, usually 10, has to be raised to in order to produce a specific number.

Logic: The study of the rules which underlie plausible reasoning in mathematics, science, law, and other disciplines.

Long odds: Poor odds, or odds which suggest an event is highly unlikely to occur.

Lottery: A contest in which entries are sold and a winner is randomly selected from the entries to receive a prize.

Mathematics: The systematic study of relationships in the physical world and relationships between symbols which need not pertain to the real world. In relation to the world, mathematics is the language of science. It operates within the laws and constraints of science as it examines physical phenomena.

Matrix: A rectangular array of variables or numbers, often shown with square brackets enclosing the array. Here "rectangular" means composed of columns of equal length, not two-dimensional. A matrix equation can represent a system of linear equations.

Median: A measure of central tendency, like an average. It is a way of describing a group of items or characteristics instead of mentioning all of them. If the items are arranged in ascending order of magnitude, the median is the value of the middle item.

Metric system: The metric system of measurement is an internationally agreed-upon set of units for expressing the amounts of various quantities such as length, mass, time, temperature, and so on.

Mode: A set of numbers is the number that occurs most frequently. There may be more than one mode. In the set (1,4,5,7), all four numbers are modes. But in the set (1,4,4,6), 4 is the only mode. The mode is one of the measures of central tendency, the others being the mean and the median.

Model: A system of theoretical ideas, information, and inferences presented as a mathematical description of an entity or characteristic.

Modulus: An operator that divides a number by another number and returns the remainder.

Mortgage: A loan made for the purpose of purchasing a house or other real property.

Nash equilibrium: A set of strategies, named after John Nash, that results in the maximum benefit of each player.

Natural numbers: The ordinary numbers, 1, 2, 3, . . . with which we count. Sometimes they are called "the counting numbers."

Negative numbers: Numbers that have a value less than zero.

Nth term: The phrase 'nth term' is used to describe any term in a sequence. The n refers to its ordered place in the sequence.

Number theory: Number theory is the study of natural, or counting numbers, including prime numbers.

Odds: A shorthand method for expressing probabilities of particular events. The probability of one particular event occurring out of six possible events would be 1 in 6, also expressed as 1:6 or in fractional form as 1/6.

Operation: A method of combining the members of a set so the result is also a member of the set. Addition, subtraction, multiplication, and division of real numbers are everyday examples of mathematical operations.

Orthogonals: In art, the diagonal lines that run from the edges of the composition to the vanishing point.

Output: Output data to a function. The output data, the dependent variable(s), that define a model.

Parabola: The open curve formed by the intersection of a plane and a right circular cone. It occurs when the plane is parallel to one of the generatrices of the cone.

Parallel: Two or more lines (or planes) are said to be parallel if they lie in the same plane (or space) and have no

point in common, no matter how far they are extended.

Parallelogram: A plane figure of four sides whose opposite sides are parallel. A rhombus is a parallelogram with all four sides of equal length; a rectangle is a parallelogram whose adjacent sides are perpendicular; and a square is a parallelogram whose adjacent sides are both perpendicular and equal in length.

Percent: From Latin for *per centum* meaning per hundred, a special type of ratio in which the second value is used to represent the amount present with respect to the whole. Expressed as a percentage, the ratio times 100 (e.g., 78/100 = .78 and so .78 × 100 = 78%).

Perfect number: A number that is equal to the sum of its divisors.

Permutations: All of the potential choices or outcomes available from any given point.

Pi: The ratio of the circumference of a circle to the diameter: $\pi = C/d$ where C is the circumference and d is the diameter. This fact was known to the ancient Egyptians who used π for the number 22/7 (3.14159) which is accurate enough for most applications.

Pixel: Short for "picture unit," a pixel is the smallest unit of a computer graphic or image. It is also represented as a binary number.

Player: In game theory, a decision maker.

Plays: In game theory, choices that can be made.

Point: A geometric element defined only by an ordered set of coordinates.

Polar angle: The angle between the line drawn from a point to the center of a circle and the *x* axis. The angle is taken by rotating counterclockwise from the *x* axis.

Polar coordinate: A two-dimensional coordinate system that is based on circular symmetry. It has two coordinates, the radius and the polar angle.

Polar-coordinate system: One of the several systems for addressing points in the plane is the polar-coordinate system. In this system a point P is identified with an ordered pair (r,θ) where r is a distance and θ an angle.

Positive numbers: Commonly defined as numbers greater than zero, the numbers to the right of zero on the number line. Zero is not a positive number. The opposite, or additive inverse, of a positive number is a negative number. Negative numbers are always preceded by a negative sign (−), while positive numbers are only preceded by a positive sign (+) when it is required to avoid confusion.

Powers: The number of times that a base is to be multiplied by itself.

Prime factorization: The process of finding all the divisors of a number that are prime numbers.

Prime number: Any number greater than 1 that can only be divided by 1 and itself.

Probability: The likelihood that a particular event will occur within a specified period of time. A branch of mathematics used to predict future events.

Probability distribution: The expected pattern of random occurrences in nature.

Probability theory: A branch of mathematics concerned with determining the long run frequency or chance that a given event will occur. This chance is determined by dividing the number of selected events by the number of total events possible.

Program: A sequence of instructions, written in a mathematical language, that accomplish a certain task.

Proper fraction: A fraction whose value is less than 1.

Proportion: Two quantities with equal ratios.

Public key system: A cryptographic algorithm that uses one key for encryption and a second key for decryption.

Pythagorean theorem: A theorem of geometry, often attributed to Pythagoras of Samos (Greece) in the sixth century B.C., states the sides a, b, and c of a right triangle satisfy the relation $c^2=a^2+b^2$ where c is the length of the hypotenuse of the triangle and a and b are the lengths of the other two sides.

Quadrilateral: A polygon with four sides. Special cases of a quadrilateral are: (1) A trapezium—A quadrilateral with no pairs of opposite sides parallel; (2) A trapezoid—A quadrilateral with one pair of sides parallel; (3) A parallelogram—A quadrilateral with two pairs of sides parallel; (4) A rectangle—A parallelogram with all angles right angles; (5) A square.

Radius: The distance from the center of a circle to its perimeter.

Rate: A comparison of the change in one quantity, such as distance, temperature, weight, or time, to the change in a second quantity of this type. The comparison is often shown as a formula, a ratio, or a fraction, dividing the change in the first quantity by the change in the second quantity. When the changes being compared occur over a measurable period of time, their ratio determines an average rate of change.

Ratio: The ratio of a to b is a way to convey the idea of relative magnitude of two amounts. Thus if the number a is always twice the number b, we can say that the ratio of a to b is "2 to 1." This ratio is sometimes written 2:1. Today, however, it is more common to write a ratio as a fraction, in this case 2/1.

Rational number: A number that can be expressed as the ratio of two integers such as 3/4 (the ration of 3 to 4) or −5:10 (the ration of −5 to 10).

Real number: Any number which can be represented by a point on a number line. The numbers 3.5, −.003, 2/3, etc. are all real numbers.

Reciprocal: The reciprocal of a number is 1 divided by the number. Thus the reciprocal of 3 is 1/3. If a number a is the reciprocal of the number b, then b is the reciprocal of a. The product of a number and its reciprocal is 1.

Reconcile: To make two accounts match; specifically, the process of making one's personal records match the latest records issued by a bank or financial institution.

Rectangle: A quadrilateral whose angles are all right angles. The opposite sides of a rectangle are parallel and equal in length. Any side can be chosen as the base and the altitude is the length of a perpendicular line segment between the base and the opposite side. A diagonal is either of the line segments joining opposite vertices.

Reflection: The operation of moving all the points to an equal distance, on the opposite side of a line of reflection.

Register: A record of spending, such as a check register, which is used to track checks written for later reconciliation.

Root: The solutions of a polynomial equation, of which the square and cube root are special cases.

Rotation: The operation of moving all the points of an object through a fixed angle around a fixed point.

Scale: The ratio of the size of an object to the size of its representation.

Scientific notation: A shorthand way to write very large or very small numbers.

Segment: A portion truncated from a geometric figure by one or more points, lines, or planes; the finite part of a line bounded by two points in the line.

Set: A collection of elements.

Simple random sampling: A sampling method that provides every element equal chance of being selected.

Statistics: Branch of mathematics devoted to the collection, compilation, display, and interpretation of numerical data. In general, the field can be divided into two major subgroups, descriptive statistics and inferential statistics. The former subject deals primarily with the accumulation and presentation of numerical data, while the latter focuses on predictions.

Stockholder: The partial owner of a public corporation, whose ownership is contained in one or more shares of stock. Also called a shareholder.

Stratified sampling: In this type of random sampling, elements are grouped together before sampling.

Symmetric key system: A cryptographic algorithm that uses the same key for encryption and decryption.

Symmetry: An object that is left unchanged by an operation has a symmetry.

Symmetry, or balance: A design is symmetrical if its two opposite sides divided by a line in the center are identical, or nearly identical.

System of equations: A group of equations that all involve the same variables.

Systematic sampling: In this type of sampling, there are intervals between each selection for sampling.

Term: A number, variable, or product of numbers and variables, separated in an equation by the signs of addition and equality.

Translation: The operation of moving each point a fixed distance in the same direction.

Trigonometry: A branch of applied mathematics concerned with the relationship between angles and their sides and the calculations based on them. First developed as a branch of geometry focusing on triangles during the third century B.C, trigonometry was used extensively for astronomical measurements. The major trigonometric functions, including sine, cosine, and tangent, were first defined as ratios of sides in a right triangle.

Unit fraction: A fraction with 1 in the numerator.

Vanishing point: In art, the place on the horizon toward which all other lines converge; a focus point.

Variable: A symbol representing a quantity that may assume any value within a predefined range.

Vector: A quantity consisting of magnitude and direction, usually represented by an arrow whose length represents the magnitude and whose orientation in space represents the direction.

Volume: The amount of space occupied by a three-dimensional object as measured in cubic units.

Whole number: Any positive number, including zero, with no fraction or decimal.

Zero: The absence of a quality (normalizing) or absence of quantity (numerical). It can also be a reference point, such as 0° on a temperature scale. In a mathematical system, zero is the additive identity. It is a number that can be added to any given number to yield a sum equal to the given number. Symbolically, it is a number 0, such that $a + 0 = a$ for any number a.

Zero-sum game: An outcome of a game where players choices have produced neither a win or a draw for all of the players.

Field of Application Index

An *italicized* page number indicates a photo, figure (f), illustration or table (t).

A

Agriculture
- Livestock sex ratios, 447
- Mendelian inheritance, 443–445
- Plant sampling, 460
- Soil tests, 410–411, 459–460, 472

Architecture
- Acoustic design, 348
- Architectural math, 33–44, 310, 590–591
- Bridge design, 76–77
- Floor numbers, 359
- Geometric shapes, 237–239
- Scale models, 468
- Skyscrapers, 19–20
- Software, 470
- Squared and cubed roots, 513
- Symmetry, 541

Arts and Crafts
- Children's books, 392
- Computer-generated, 22
- Creativity, 285
- Drawings and paintings, 267, 391–392
- Fractal patterns, 22, 201
- Illustration, 392
- Jewelry, 42
- Photographic negatives, 281
- Proportion, *433*, 433–434
- Size and shape, 21–22

Astronomy and Space Travel
- Age of the Universe, 160
- Antenna design, 535–536
- Astronomical averages, 55
- Computational astronomy, 487
- Cosmology, 487
- Expanding Universe, 178–179
- Galaxy arrangement, 202
- Geostatistics, 525–526
- Light years, 486–487
- Mars, 462–463
- Moon, 447–448, 451, 462–463
- Rocket launches, 480–481
- Solar systems, 242–243, 471
- Space navigation, 93
- Space Shuttle steering, 305
- Space-time continuum, 245–247
- Space travel, 24, 311, 312, 535
- Speed of light, 486–487
- Star magnitude, 418
- Telescopes, 43, 165–166
- Timekeeping in space, 312

B

Business
- Accounting, 357–358
- Bankruptcy, 150
- Burn rate, 151
- Business math, 62–68
- Business plans, 591
- Cash registers, 73
- Charts, 111–113
- Check processing, *291*, 292
- Compound interest, 74–75, 376
- Credit card fraud, 27–28
- Currency conversion, 129
- Demographics, 143
- Distribution of earnings, 150–151
- E-money, 160
- Factoring, 182–183
- Fundraising, 19
- Game theory, 228, 229
- Global economics, 515
- Insurance averages, 57
- Just in time manufacturing, 440
- Marketing and sales, 30, 341, 375, 463, 589
- Online auctions, 230
- Organizational charts, 413
- Pilfering, 153–154
- Product samples, 463–464
- Production management charts, 413
- Profits, 86–87, 531–532, 588
- Project management, 285
- Proportion and inverse proportion, 586–587
- Public opinion polls, 379
- Quality control, 526–527, 592–593
- Simple averages, 152
- Staffing levels, 589
- Stock levels (Goods), 592
- Stock market, 408–409, 436, 515, 597
- Store assistants, 592
- Time management, 413
- Weighted averages, 54–55

C

Communications
- Air traffic control, 135
- Antennas, 202, 535–536
- Data transmission, 119–120
- Information theory, 273–274
- Internet, 29–30, 117–118
- Language translation, 587
- Linguistic algorithms, 31
- Submarines, 282
- Subnet mask, 118

Computers
- Aircraft control, 157
- Algorithms, 27
- Animation, 134, 392–393
- Artificial intelligence, 32, 230
- Base numbering systems, 60–61
- Binary logic, 417–418
- Boolean algebra, 146, 301
- Charts, 112
- Computer intensive applications, 297
- Computer monitors, 117, 427–428

Computers and math, 114–121
Data compression and transmission, 28, 87–89, 118–120, 202, 264, 349
Digital images, 29, 164–165, 263–264, 304–305, 403, 470–471
Domain and range of functions, 157, 158
Error-correction codes, 275–276, 349, 363–364
File sorting systems, 425
Fractal compression, 202
Fuzzy logic, 301–302
Graphics, 470–471, 566, 572, 581
Internal organization, 412
Internet, 29–30, 117–118
Linear programming, 291–292
Macros, 286
Memory and speed, 310
Music, 349–350
Networks, 259–260, 555
Quantum computing, 276–277
Scientific notation, 487–488
Software, 166, 285, 584
Subnet mask, 118
Supercomputers, 78
Web search engines, 146
Word problem programming, 584
Construction
Building materials, 579
Carpentry, 37, 479–480
Electric circuit diagrams, 414
Fraction measurements, 208
Pothole covers, 236–237
Surveying, *49*, 49–50

E

Education
Comparisons, 586
Demographics, 112–113
Grades, 54, 139–140, 140*f*, 153, 452–456
Gumball estimation, 163–164
Math tables, 544–545
Student loan consolidation, 56–57
Student-teacher ratio, 445
Task-specific rubrics, 455–456
Testing, 382–383, 555, 589–590
Time management skills, 153
Word problems, 584
Energy
Municipal electricity supply, 589
Nuclear, 206, 213–214, 311
Oil and gas, 258, 451
Prediction of demand, 93–95, 94*f*
Solar, 50
Sonoluminescence, 579–580
Engineering
Automobile design, 305–306, 435
Bridges, 76–77
Designing for strength, 89
Domain and range of functions, 157–158
Electric circuits, 486
Electronic engineering, 489
Factoring, 182
Failure prediction, 89–90
Fluid mechanics, 281
Machine parts, 579
Measurement systems, 309
Nonlinear design, 281
Quality assurance, 526–527
Scientific notation, 487
Ships and submarines, 482
Vehicle weight, 535
Well depth, 594
Word problems, 585–586
Entertainment
Auditorium and concert hall acoustics, 348
Countdowns, 533
Dance, 266
Fireworks, 241
Games and puzzles, 134, 220, 223, 228–229
Magic, 221–223, 428–429
Möbius strip, 555–556
Movies, 28, 392–393, 393–394
Music, 285, 292–293, 347–348, 348–349, 349–350, 351–352, 435, 445, 471
Rhythm, 206–207
Rubik's Cube, *243*, 243–244
Sound, 213, 241–242, 282–283, 297
Environment. *See* Nature and Wildlife

G

Games. *See* Sports and Recreation
Geography and Mapping
Cartography, 100–106
Conformal maps, 159–160
Continuously operating reference station (CORS), 565
Coordinate systems, 135
Demographics, 143
Geodesy, 564
Global positioning systems, 105, 239–241, 258, 565
Great Trigonomic Survey, 564–565
Land area conversions, 129
Map diagrams, 413–414
Map scale, 434, 467–468
Mapping algorithms, 30
Mount Everest height, 310
Oil exploration, 258
Paper maps, 134–135
Surveying, 136, 258, 564, 592
Government and Politics
Average families, 55–56
Census, 142–143, 462
City planning, 42
Currency design, 211–212
Election demographics, 141–142
Government salaries, 548
Mailman's route, 589, 589*f*
Municipal electricity, 589
National debt, 5
News media statistics, 378–379, 380
Population statistics, 451
Public opinion polls, 379, 527–528, 593
State lottery odds, 207–208, 368–369
Voting, 208–209
War, 597

H

Health and Medicine
Alcohol-caused brain damage, 578
Alzheimer's disease, 578
Arm span and height, 408
Autism, 594
Bacterial growth, 326–327, 481
Biomedical graphing, 258
Blood pressure measurement, 310
Blood tests, 472
Body composition, 578–579
Body diagrams, 414
Body mass index, 214, 319
Brain volume, 578, 581–582
Cancer treatment logarithms, 298–299
CAT scans, 147, 480
Causality and disease, 585
Cigarette smoking, 4
Clinical trials, 323–326, 460, 461, 488
Diagnosis, 264, 318–319, 320*t*, 480
Diet and fitness, 3–4, 259, 436, 534, 548–549, 548*f*, 549*f*
Disease demographics, 143, 230
Disease outbreaks, 591
Disease probabilities, 208
Drugs and pharmaceuticals, 47–48, 147, 310–311, 434, 460–461
Ergonomics, 42, 434–435
Food inspection imaging, 266
Genetics, 28, 30, 274–275, 321–323, 443–445, 482, 483, 493, 545
Height and weight charts, 319–320, 446, 593
Hospital size, 440
Human body proportions, 436–437
Infectious disease modeling, 230
Life expectancy, 57, 549, 550*f*
Polychlorinated aromatic hydrocarbons, 578
Prognosis, 369
Proteins, 490
Robotic surgery, 245
Sleep management, 534
Sound intensity measurement, 297

Stem cell research, 446–447
Vision inversion, 281
Visualization software, 258
Water safety, 446
Household Applications
Area of shapes, 340–341, 439
Automobile insurance, 428
Bed size, 435
Calendars, 98–99, 469
Decision making, 597
Drinking water, 446
Fabrics, 43
Food and cooking, 127, 129, 173–174, 206, 302, 446, 591
Instructional diagrams, 414–415
Interior design, 394–395, 434, 594
Landscaping, 386, 396
Length measurements, 450
New Year's resolutions, 534
News media statistics, 378–379, 380
Packing objects, 590
Swimming pools, 446, 581
Time measurement, 311, 452
Toys, 471
Volume calculations, 439
Weight measurements, 450, 468–469
Work schedules, 549

M

Maps. *See* Geography and Mapping

N

Nature and Wildlife
Animals, 147–148, 164, 421–422, 587
Atmospheric pressure, 469
Biomechanics, 354
Earth science, 446, 489
Environment, 418, 462
Evolution, 57–58, 174–175
Fractal patterns, 200–201, 541–542
Game theory, 229–230
Global warming, 48–49, 257–258, 594
Honeycombs, 239, *240*
Modeling, 354–355
Molecular cloning, 286
Natural disasters, 289, 358–359, 417, 471, 482–483, 591–592
Natural resources evaluation, 105
Nature and numbers, 353–355
Ocean level changes, 580
Patterns, 286, 353–354, 355
Photography, 402–403
Phytoplankton, 354–355
Plant sampling, 460
Population, 171–173, 174–175, 305, 338, 353–354, 408, 409*f*
River flow, 488
Water runoff, 582
Weather, 77–78, 126–127, 201, 414, 461, 476–478, 593
Wind strength, 469

P

Personal Finance
ATMs, 73
Automobile purchases, 187
Budgeting, 151, 188–198, 549–550
Cash registers, 73
Cellular service plans, 586
Checkbook, 13–14, 189–190, 546
Compound interest, 376, 591
Credit cards, 27–28, 178, 185–186, 369
Currency exchange rates, *195*, 195–196, 586
Currency trading, 597
E-money, 160
Futures trading, 597
Income taxes, 14, 31–32, 189, 532, 548
Inflation exponentials, 177–178
Insurance, 584–585
Interest rates, 177–178, 546–547, 547*f*, 591
Investments, 190, 337–338, 546–548
Loan rates, 547–547
Lottery probability, 207–208
Mortgage rates, 547–547
Music purchases, 184–185
Overtime pay, 208
Pay rates, 339
Retirement plans, 191–194
Savings per purchase, 341
Social Security system, 190
Value added taxes, 592

S

Science and Technology
Absolute dating, 489
Aerodynamics and hydrodynamics, 259
Archeology, 27, 310, 463, 585
Area between curves, 91–96
Carbon dating, 165, 298
Chemicals, 311, 435–436, 486, 566
Combinatorial chemistry, 147
Domain and range of functions, 157–158
Failure prediction, 56, 89–90, 427–428
Filtering, 48
Finite-element models, 212–213
Fireworks, 241
Forensic science, 488–489
Fossil samples, 460
Geologic time scale, 488
Gravity, 313
Guidance systems, 91–96, 93*f*, 440
Historical conversions, 125
Human motion tracking, 290–291
Hypothesis testing, 585
Land mine detection, 572–573
Mass conversions, 129
Measurement systems, 139
Military models, 331–332
Nanotechnology, 418, 490
Nuclear waste half-life, 206, 213–214
Optics, 264, 298
Periodic table, 545
pH scale, 418
Physics and information theory, 274
Precision measurements, 452
Radar systems, 136–137
Radiation shielding, 299
Radioactive dating, 177
Radioactive decay, 175–177, *176*, 176*f*
Random number generators, 75–76
Reducing equations, 181–182
Riemann hypothesis, 212
Scientific notation, 171
Sound reduction, 282
Spectra, 90–91
Stealth technology, 244–245
Temperature measurements, 125–126, 127, 357
Thermometers, 580–581
Time measurement, 311
Volume conversions, 129
Security
Biometrics, 22–23
Closed circuit television, 31
Codes and code breaking, 182
Credit card security, 147
Cryptography, 28, 147, 158, 280–281, 299, 585
Digital watermarks, 266–267
E-money, 160
Encryption, 28–29, 120–121, 362–363, 424–426
Face recognition software, 264–266
Forensics, 266, 425
Group theory, 299
Military perimeters, 387–388
Perimeter barriers, 386
Random number generators, 75–76
Shopping and Consumers
Automobile purchases, 153, 162–163, 187
Bulk purchases, 450
Buying by area, 48
Cellular service plans, 187–188
Check-out line speed, 367–368
Consumer affluence, 153
Credit card security, 147
Decision making, 181
Demographics, 143
Jewelry pricing, 573–574
Lightbulbs, 20–21
Marketing and sales analysis, 30
Pricing, 4, 376–377, 451–452, 586

Shopping and Consumers (*cont.*)
Product samples, 463–464
Rebates, 377, 532–533
Sales tax, 377
Savings per purchase, 341
Tip calculations, 194–195, 375–376
Warranty periods, 427–428
Sports and Recreation
ACL injuries, 573
Archery, 244
Athletic performance, 152, 154, 497–498
Attendance statistics, 409
Baseball, 53–54, 139, 152, 338–339, 381–382, 426, 498–499, 504–505
Basketball, 381, 445–446, 498, 499–500, 503, 507
Bicycling, 505–506
Bowling, 3, 426
Card games, 5, 77, 218–219, 366
Casino games, 219–220
Championship standings, 31, 382
Countdowns, 533
Downhill skiing, 339
Drag racing, 572
Electronic timing, 339
Field and court design, 40–41
Figure skating, 504
Football, 14–15, 182, 358, 382, 386, 499, 500–502, 507–508
Gambling and betting, 6, 366–367, 425–426, 508–510, 591
Golf, 358, 506–507, 533–534
Horse racing, 339
Ice hockey, 498
Lottery odds, 207–208, 368–369, 591
Media experts, 502–503
Movement technique, 573
Photography, 402–403
Playing field surveying, 514–515
Rally racing, 586
Ratings percentage index, 503–504
Repetition, 284–285
Scoring systems, 3
Skydiving, 17–18
Soccer, 506
Space tourism, 24
Speed measurements, 311
Sports math, 495–510
Statistics, *380*, 380–382
Swimming pools, 386–387, 581
Throwing a ball, 593
Timing games, 139
Tournaments, 590
Track and field, 309, 358, 534
Video analysis, 263
Virtual reality games, 412
Virtual tennis, 291
Winning percentages, 154
Zero-sum games, 596

T

Transportation
Aerodynamics and hydrodynamics, 259
Aircraft and airplanes, 16–17, 135, 157, 211, 429
Automobiles, 49, 153, 162–163, 187, 305–306, 408, 428, 435, 439–440, 445
Crash tests, 18–19
Engine compression, 579
Friction and weight, 563–564
Radar systems, 136–137
Vehicle weight, 451, 535
Wheels, 43
Travel
Automobile GPS systems, 470
Distance measurements, 128, 308–309, 452
Ferry schedules, 563
Flying *vs.* driving, 4, 370
Gasoline cost per trip, 443
Greenwich Mean Time, 135
Metric conversion for, 339–340
Miles per gallon calculations, 341, 452
Navigation, 73–74, 311–312, 514, 562–563, 587–588, 592
Schedules, 549, *551*
Shortest distance, 587–588
Space tourism, 24
Street signs, 414
Supersonic and hypersonic flight, 299
Teleportation, 23–24
Time estimation, 586
Time travel, 246–247
Time zones, 547
Toll roads, 533
Trip length, 443

General Index

A **boldface** page number indicates the main essay for a topic. An *italicized* page number indicates a photo, figure (f), illustration, or table (t).

1% method, 374
80/20 rule, 370–371
128-bit encryption, 425
365-day calendars, 98
401(k) plans, 191–192
1099 Form, 189

A

AAD (Automatic activation device), 18
Abacus, 1–2, *2*, 71–72, 74
Abelian class field theory, 361
Abscissa, 254–257
Absolute dating, 489
Absolute temperature, 127
Absolute zero, 127, 356, 466–467
Abstract symmetry, 542
Acceleration, 83, 307, 439, 572
Accidents
  aircraft, 429
  automobile, 5, 18–19
  mortality from, 5
Account numbers, credit card, 369
Accountants, 63–64
Accounting, 63–65, 357–358, 451–452
Accuracy, 55
Acid rain, 418
Acids, 418
ACL (Anterior cruciate ligament), 573
Acoustic design (Architecture), 348
Acoustic instruments, 242
Acquired immune deficiency syndrome (AIDS), 322, 461
ACT, dense-dose, 323–325
Actuarial tables, 428
Actuaries, 584–585
Acute angles, 558
Acute triangles, 559
Adding machines, 2
Addition, **1–8**
  algorithms, 27
  notations, 485
  tables, 544, 544*f*
  of vectors, 570–571
Adleman, Leonard, 28, 362
Adolescents, 578
Aerodynamics, 259
Affluence, consumer, 153
Aggression, 229–230
AIDS (Acquired immune deficiency syndrome), 322, 461
Air pollution, 462
Air traffic control, 131, 135, *570*
Air travel
  around the world, 481
  vs. driving, 4, 370
  flight insurance for, 370
Airbags, 18
Aircraft
  accidents, 429
  Bernoulli's equation and, 211
  computer systems in, 157
  design, 16–17
  flight mathematics, 479
  fuel consumption, 16–17
  guidance systems, 440
  rain on, 474, 479
  stealth technology, 244–245
  supersonic and hypersonic, 299
  take-offs and landings, 478–479
  weight of, 535
Aircraft carriers, 478
Alarm clocks, 161–162
Albers equal area conic projection, **102**, **102**
Alberti, Leon Battista, 390–391
Alcohol, 578
Alexander Nevsky (Film), 349
Algebra, **9–25**
  Boolean, 145, 146, 301, 302
  development of, 12–13
  fractions and, 205
  fundamental concepts, 9–12
  linear, 571
  logarithms and, 296
  matrix, 146, 262, 303, 304, 304*f*, 572
  potential, 23–24
  powers and, 296
  real-life applications, 13–23
  vector, 570–571, 571*f*
Algebraic number theory, 361
Algorithms, **26–32**
  backtracking, 27
  coding theory and, 119
  computers and, 115
  data transmission, 119–120
  De la Loubere's, 222
  Diffie Hellman key agreement, 362
  digital signal, 363
  Dijkstra's, 588
  discrete math for, 145–146
  El Gamal, 363
  Euclidean, 361
  Floyd's, 588
  Kruskal's, 589
  logical, 27
  MD5, 425
  repetition, 119
  RSA, 28, 362, 363
  seeding, 31, *31*
  SKIPJACK, 29
  sorting, 590
  transposition-substitution, 29
  XOR, 362
Almagest, 562
ALU (Arithmetic logic unit), 3
Aluminum bats, 504–505
Alzheimer's disease, 578
Ampere, 123

Amplification (Sound), 347–348
Analog recordings, 348
Analysis of variance (ANOVA), 522–523
Analytic geometry, 13
Analytic number theory, 361
Analytic rubrics, 455
Analytical models, 329
Andromeda Galaxy, 486
Anemia, sickle-cell, 321
Angles
  acute, 558
  degrees of, 71, 557, 558
  measurement of, 308, 557–558
  obtuse, 558
  plane, 123, 557–558
  right, 558
  sine function of, 560–561, 560*f,* 562
  solid, 123
Animals
  behavior, 229–230
  breeding, 483
  See also Population growth; Wildlife
Animation
  coordinate systems for, 134
  digital, 28
  perspective in, 392–393
  three-dimensional, 263
ANOVA (Analysis of variance), 522–523
Antennas
  cellular telephone service, 202
  radio, 182, 202
  satellite, 536
  space exploration and, 535–536
Anterior cruciate ligament (ACL), 573
Anthropology, 489
Anti-sound, 182
Anytime minutes, 187
Apartments, 384, 433
Aperture, lens, 400, 401, 402–403
Apollo mission, 535
Approximation, Diophantine, 361
Arch of Constantine, *39,* 40
Archeology, 27, 310, 463, 585
Archery, 244
Arches, semicircular, 237
Archimedes, 162, 385, 442, 577
Archimedes principle, 482
Architectural math, **33–44**
  acoustic design, 348
  development of, 35–36
  floor numbering, 359
  fundamental concepts, 33–34
  geometry and, 237–239, *239*
  Golden ratio in, 443, 492
  measurement systems, 35, 310
  perspective in, 389–391, 396
  proportion and, 34, *34,* 38, 432–433
  real-life applications, 36–43
  scale in, 467, 468
  software for, 470
  squared and cubed roots in, 513–514
  symmetry in, 34, *35,* 36–38, *39,* 541, *541*
  word problems, 590–591
Architectural models, 433, 468
Area, **45–50**, 46*t*
  measurement conversions for, 124
  multiplication for, 340–341
  quadratic and cubic equations for, 439
  surface, *47,* 48, 439, 576
Area graphs, 252, 253*f,* 405
Aristotle, 53, 300
Arithmetic logic unit (ALU), 3
Arithmetic mean, 51–52, 519
Arithmetica (Diophantus), 12
Arm span, 408
Aromatase inhibitors, 325–326
Arrays, 475
Arron, Hank, 517
Arrows, 244
  See also Vectors
Ars Nova, 346
Art
  computer-generated, *21*
  digital images of, 267
  perspective, 391–392, 393
  proportion, 432, 433–434
Artificial intelligence, 32, 230
Asset division, 150
Astronomical charts, 297
Astronomical tables, 544
Astronomy
  astronomical charts and tables, 297, 544
  averages and, 53, 55
  computational, 487
  data transmission for, 120
  domain and range of functions, 157
  fractals in, 202
  music and, 346
  telescopes, 43, 55, 165–166
  trigonometry and, 562
Asymmetric encryption. *See* Public key encryption
Athletic performance, 152, 154, 445–446, 495, 497–498
  See also Sports math
Atlas Eclipticalis (Cage), 349
ATM (Automatic teller machine), 73
Atmospheric perspective, 391
Atmospheric pressure, 469
Attendance counters, 4–5
Auctions, online, 230, 451, 531–532
Audio compression, 119
Auditoriums, 348
Authentication, data, 120
Autism, 594
AutoCAD, 470
Automatic activation device (AAD), 18
Automatic teller machine (ATM), 73
Automobile accidents, 5, 18–19
Automobile insurance, 57, 428
Automobiles
  acceleration, 307, 439
  vs. air travel, 4, 370
  braking distance, 408
  buying, 153, 162–163, 187
  computer-aided design, 305–306
  crash tests, 18–19
  engine compression ratio, 445, 579
  ergonomic design, 42
  fuel consumption, 341, 352, 383, 451
  GPS systems, 470
  loans, 547
  models of, 435
  monthly payments, 187
  oil change, 154–155
  performance percentages, 383
  price, 4
  radiators, 49
  revolutions per minute, 445
  seat design, 435
  tires, 439–440
  trade-ins, 187
  vs. trucks, 441–442
  used, 162–163
  weight of, 535
  wheels, 43
Average, **51–58**, *54*
  deviation, 520
  division for, 152–153
  grade point, 139–140, 140*f,* 153
  moving, 259
  statistical, 519
  weighted, 54–55, 153
Avogadro's law, 486
Axis, 131, 248–249, 404
Azimuths, 562–563

B

Babies
  development, 343
  products for, 143
Baby boom, 143
Babylonian Talmud, 226, 228
Babylonians
  calendars, 558
  counting boards, 71
  degrees of angle, 557–558
  geometry, 235–236
  multiplication tables, 335–336
  pi, 442
  powers, 416–417
  tables, 543–544
  weight measurements, 308
Bach, Sebastian, 347
Backtracking algorithms, 27
Bacteria
  game theory modeling and, 230
  population growth, 173–174, 326–327, 326*f,* 481

Bakine, 127, *127*
Balanced diet, 436
Balloons, weather, 477–478
Ballot boxes, *7*
Balls
  area of, 46
  basketballs, 75
  throwing, 593
  See also Spheres
Bank check processing, *291,* 292
Bankruptcy, 150
Bar charts, 109–110, 110*f,* 111, 111*f,* 112
Bar graphs, 249–251, 250*f,* 251*f,* 406, 410, 411*f*
Bar scale, 468
Barcodes, *15,* 15–16, *23*
Barometers, 469
Barriers, perimeter, 386, 387–388
Barter, 205
Base, **59–61,** 167, 295–296
Base 2 numbering systems. *See* Binary numbering system
Base 5 numbering systems, 59–60, 60*f*
Base 8 numbering systems, 60, 61
Base 10 numbering systems. *See* Decimals
Base 20 numbering systems, 60
Base 60 numbering systems, 60, 557–558
Base e, 297
Base rate, 374–375
Base sequences. *See* Genetic sequence
Baseball
  bats, 504–505
  batting averages, 53–54, 152, 381–382, 426, 498, 517
  commentators, 502
  earned run average, 338–339, 498
  on-base percentage, 498
  runs scored, 498
  speed of pitches, 139
  statistics, 498–499
  World Series wins, 427
Bases (Chemistry), 418
Basketball
  championship standings, 382
  commentators, 502
  free-throws, 500
  game theory and, 225
  iteration in, 284
  most valuable player, 381
  player's height, 368
  point average, 498
  point guard performance, 445–446
  Ratings percentage index, 503
  salary caps, 507
  shooting percentages, 374
  statistics, 499–500
Basketballs, size of, 75
Batello division, 149–150
Bats, baseball, 504–505
Batting average, 53–54, 152, 381–382, 426, 498, 517
Bayer array, 401
Bayes, Thomas, 429
Beak size, 58
Beams, 395
Bearing (Navigation), 562–563
Bears, polar, 576
Beaufort, Francis, 469
Beaufort scale, 469
A Beautiful Mind, (Film) 228
Beauty and the Beast, (Film) 393
Beauty principle, 198
Beckham, David, 506
Beds, 42, 435, 440
Bees, honey, 239
Beethoven, Ludwig van, 432, 443
Behavior
  animal, 229–230
  game theory and, 225–231
  patterns of, 181
Bernoulli, Daniel, 211
Bernoulli's equation, 211
Best fit, line of, 404
Betting, 508–510. *See* Gambling
Betto Bardi, Donato di Niccolo di, 391
Bicycles, 505–506
Big bang, 178, 487
Bilateral symmetry, 38
Binary numbering system
  calculators and, 70
  computers and, 60, 114–115, 116–117, *117*
  Egyptian, 336
  index searches, 590
  powers and, 417–418
Binomial distribution, 317, 317*f,* 518
Biomechanics, 354–355
Biomedical research graphs, 258
Biometrics, 22–23, 581–582
Birth rates, 525
Birthday predictions, 428–429
Bits, 116, 271–272
Black and white photographs, 281
Black powder, 241
Blackjack, 217, 218
Blood pressure measurement, 310
Blood tests, 472
Blue screen images, 134
BMI (Body mass index), 214, 319
Board games, 134, 220, *220*
Body, human. *See* Human body
Body mass index (BMI), 214, 319
Body size (Wildlife), 576
Body surface area (BSA), 48
Boethius, Anicis Manlius Severinus, 345
Boiling point, 126
Bokeh, 403
Bolts, 526–527
Bonds, 67–68, 191
Bookkeeping, 63–65, 357–358, 451–452
Books
  children's, 392
  comic, 392
Boole, George, 145, 146, 301
Boolean algebra, 145, 146, 301, 302
Borders, geographic, 234
Borrowing (Subtraction), 531
Boston Red Sox, 427
Boundaries, perimeter, 385–388
Bowl Championship Services, 14–15
Bowling, 3, 426
Box plots, 404, 407, 408*f*
Boxes, 46, 575
Bradley, General, 332–335
Brahe, Tycho, 346
Brahmagupta, 357
Braided rope, 539, 539*f*
Brain and vision, 281
Brain volume, 578, 581–582
Braking distance, 408
Break-even point, 184–185
Breast cancer, 323–326, 324*f,* 327
Breeding animals, 483
Bridges
  derivatives for, 89
  design, 76–77, *90*
  failure prediction, 89–90, *90*
  span, 478
Brilouin, L., 274
British pound, 196
Brown, Robert, 350
Brown noise, 350
Brownell, William, 531
Brunellesci, Filippo Di Ser, 389–390, 390*f*
BSA (Body surface area), 48
Bubble fusion, 579–580
Bubble graphs, 257, 257*f*
Budgets
  business, 63, 64*f*
  division for, 151
  food, 194
  personal, 188–189
Building materials, 579
Bulk purchasing, 450
Bull riders, 339
Bürgi, Joost, 72
Burn rate, 151
Burst errors, 275
Business math, **62–68**, *65, 67,* 182–183
Business plans, 591
Butterfly effect, 199
Bytes, 116, 118

C

C note, 344
CAD (Computer-aided design), 305–306
Caesar, Julius, 98
Cage, John, 349

Calculators, **69–79**
development of, 2, 71–73, *72,* 115–116
fundamental concepts, 69–71
graphing, 69, *70,* 257
mechanical, 72
potential applications, 78
real-life applications, 73–78
scientific, 69
software as, 72–73
speed-distance-time, 74
for trigonometry, 564
Calculus, **80–96**
area and, 46
development of, 86, 87
domain and range, 156
fundamental concepts, 81–86
integration, 46, 577
real-life applications, 86–95
Calendars, **97–99**
365-day, 98, 99
Babylonian, 558
base and, 59
Gregorian, 98, 99
Julian, 98, 99, 451
lunar, 97–98, *98,* 99
lunisolar, 98, 99
solar, 98
as time scales, 469
Calorie counting, 3–4, *320,* 436, 534
Camera lenses, 399, *399,* 400*f,* 402–403
Camera obscura, 391, 402
Cameras, 398–403
development of, 402
digital, 29, 154, 401
Kodak, 402
shutter speed, 399, 401
Canada
currency exchange rates, 196
travel to, 339–340
Cancer
breast, 323–326, 324*f,* 327
radiation therapy, 298–299
screening, 320*t*
Candela, 123
Candlestick plots, 259
Capacitors, 489
Capital investments, 63, 66–67
Car seats, 42
Carats, 129
Carbon dating, 165, 298
Carbon dioxide, 49
Card games
combinatorics for, 77
math for, 217, 218–219
odds in, 365, 366
poker, 5, 219, 366, 596
probability theory, 5
war, 595
Cardano, Gerolano, 280, 424
Carpentry, *36,* 37, 479–480, 480*f*
Carrying (Subtraction), 530–531
Cars. *See* Automobiles
Cartesian coordinate system, 132–133, 133*f,* 134, 136, 257
Cartography, **100–106**
See also Maps
Cash registers, 73, 451–452
Casino games, 219–220, 366–367, 425–426
See also Gambling
Casino industry, 223
Cassini space probe, 93
CAT scans, 147, 475, 480
Cattle, 447
Causation, 524
CBD (Central business district), 42
CCD (Charged coupled devices), 29
CCDS (Charged-coupled device sensors), 401
CCTV (Closed circuit television), 31
CDs. *See* Compact disks
CDs (Certificates of deposit), 546–547
Cellular telephones
antennas, 202
coverage area, 188
opinion polls and, 527
selecting, 187–188, 586, 586*f*
Celsius, 123, 125–126, 340, 357, 466
Celsius, Anders, 126
Census, 142–143, 451, 462
Census Bureau (U.S.), 193, 451
Centimeters, square, 45–46
Central business district (CBD), 42
Central tendency, 52
Centric perspective. *See* Perspective
Certificates of deposit (CDs), 546–547
Challenger Space Shuttle, 56, 427
Championship standings, 31, *31,* 382, 503–504
Chance, 215–216, 316, 424
See also Odds; Probability
Channels, noisy, 274
Chants, Gregorian, 345
Chaos theory, 199, 330, 331
Charged coupled devices (CCD), 29
Charged-coupled device sensors (CCDS), 401
Charles, Jacques, 127
Charts, **107–113,** *108*
astronomical, 297
bar, 109–110, 110*f,* 111, 111*f,* 112
column, 109, 109*f,* 110, 110*f,* 111, 112, 112*f*
domain and range of functions, 158
flow, 411–412, 413*f*
height and weight, 319–320
line, 107–109, 108*f,* 111
organizational, 413
pie, 110–112, 111*f,* 113, 113*f*
polar, 406–407
religious, 297
run, 409–410
See also Graphs
Check cards, 73
Check processing, *291,* 292
Checkbooks, 13–14, 189–190, 358, 546
Checkout lines, 367–368
Checkout systems, code-scanning, 62
Chemical solutions, 435–436
Chemicals, 151–152, 566
Chemistry
combinatorial, 147
measurement systems, 311
scientific notation for, 486
Chess, 134, 216–217, *217,* 221, 226
Child development, 343
Children's books, 392
China
calendars, 99
card games, 217
colored rods, 356
Chisenbop, 2
Chlorine, 152, 446
Choking, 5
Cholera, 322
Chords, 561–562
Chronic diseases, 322
Chuquet, Nicholas, 417
Churches, 237
Cicadas, 421–422
Cigarette smoking, 4
Ciphers. *See* Cryptography; Encryption
Circles
area of, 46, *234,* 340
circumference, 385, 442, 561–562
definition, 233
properties, 237
Circuit diagrams, 414
Circumference
circle, 385, 442, 561–562
of the Earth, 105
City planning, 42, 396
Classification schemes, 493
Clerk-Maxwell, James, 402
Climate
change in, 48–49, 257–258, 580, 594
cold *vs.* hot, 576
Climbing rope, 537
Clinical trials
breast cancer, 323–326, 324*f,* 327
medical mathematics and, 314, 315
sampling for, 460, 461
scientific notation for, 488
Clocks
alarm, 161–162
countdown, 533, *533,* 535
digital, 139
Doomsday, 535
navigation and, 135
pendulum, 311
quartz, 311
rounding off time, 452
Cloning, molecular, 286
Closed circuit television (CCTV), 31
Clothing, 384
Cloud activity models, 77

Cloud area, 48–49
Cluster sampling, 458, 460, 461
Clustered column charts, 110, 110*f*
CMOS (Complementary metal oxide semiconductor sensors), 401
Coaches, 152
Coastline, Maine, 198
Cobb, Ty, 53
Cocks, Clifford, 281
Code of points system, 504
Codes (Computer)
  error-correction, 269, 275–276, 349, 360
  Pretty Good Privacy, 281
  Reed-Solomon, 276
Codes (Secret). *See* Cryptography
Code-scanning checkout systems, 62
Coding theory, 119
Coefficient of determination, 404, 405
Coefficient of restitution, 504–505, 506–507
Coefficient of the term, 11
Coefficients (Multiplication), 335
Coin toss, 279, 474, 500–501, 501*f*, 517–518
Cold climate, 576
College graduates, 524
Colored rod system, 356
Colors, mixing, 22
Colossus device, 29
Columbia Space Shuttle, 56
Columbus, Christopher, 133
Column charts, 109, 109*f*, 110, 110*f*, 111, 112, 112*f*
Combinatorial chemistry, 147
Combinatorics, 77, 145
Comic books, 392
Commentators, 502–503
Commodity futures, 191
Common terms, 204
Communication
  information theory and, 273–274
  IP address for, 117–118
  submarine technology for, 182
Commutative property, 335
Compact disks (CDs)
  data compression, 118, 349
  data transmission algorithms, 120
  design, 348
  error-correction codes, 349, 364
  storing data on, 88*f*, 89
Comparisons, 380, 586
Competition, 175
Complementary metal oxide semiconductor sensors (CMOS), 401
Completion percentage, 499
Composite line graphs, 409, 409*f*
Composite numbers, 180
Compound interest, 74–75, 170, 376, 591
Compression
  data, 118–119, 202, 264, 349
  engine, 445, 579
  sound, 349
Compulsive gamblers, 367
Computational astronomy, 487
Computed tomography (CAT scan), 147, 475, 480
Computer games
  chess, 217
  math for, 223
  perspective in, 396
  virtual reality and, 412
Computer graphics, 396, 566, 572
  See also Digital images
Computer intensive operations, 297
Computer languages. *See* Programming (Computer)
Computer monitors, 117, 427–428
Computer networks, 259–260, 555
Computer searches, 302
Computer-aided design (CAD), 305–306
Computer-generated art, *21*
Computer-generated music, 349–350
Computers, **114–121**
  arithmetic logic unit, 3
  binary numbering system and, 60, 114–115, 116–117, *117*
  business math and, 62
  calculator programs for, 72–73
  for calculus, 86
  development of, 115–116
  domain and range of functions, 157, 158
  electronic music on, 348
  fundamental concepts, 114–115
  infinity and, 150
  for language translation, 587
  macros for, 286
  memory, 310
  quantum, 276–277
  real-life applications, 116–121
  sales graphs for, 253, 253*f*
  scientific notation for, 487–488
  speed, 310
  storing data on, 87–89, 88*f*
Computing, forensic, 425
Concert halls, 241–242, 348
Concerts, rock, 297
Conclusions, 300–301, 301*t*
Concrete, 579
Coneflowers, 354
Confidence intervals, 522
Conflict, 225
Conformal maps, 159–160
Congressional Representatives, 209
Conic projections, 102
Constants, 10–11
Construction, 37, 208
Consumer affluence, 153
Consumer Price Indexes (CPI), 378–379
Continued fractions, 206
Continuous functions, 255
Continuous objects, 144
Continuously operating reference station (CORS), 565
Convenience sampling, 461
Convergent sequences, 493, 588
Convergent series, 493–494
Conversions, **122–130,** 477
  distance, 123, 128, 348, 545, 546*f*
  historical, 125
  tables for, 545
Cooking
  measurement conversions, 129
  oven temperature, 127, *127*, 302
  recipes, 206, 446, 591
  temperature conversion, 127, *127*
Coordinate systems, **131–137**
  Cartesian, 132–133, 133*f*, 134, 136, 257
  for graphs, 248–249
  for maps, 103–104
  polar, 132, 133, 133*f*, 136–137
Coordinated Universal Time (UTC), 547, 547*f*
Coordinates, 131
Copernicus, Nicolas, 346
Copyright, 267
Coriolis force, 480–481
Coriolis, Gustave, 481
Corporate earnings, 150–151
Correlation, 521–522, 524, 525–526
Correlation coefficient, 404–406, 521–522
CORS (Continuously operating reference station), 565
Cosecant, 564
Cosine, 560–561
Cosmology, 487
Cost algorithms, 588
Cost minimization, 588
Cotangent, 564
Coulomb, 124
Countdowns, 533, *533*, 535
Counterfeit money, 211–212
Counting
  base and, 59
  cards, 218–219
  in epidemiology, 322
  probability theory for, 147–148
  in sports, 495–496
  tools, 1–3, 4–5
  See also Combinatorics
Counting boards, 71, 116
Counting tables (Abacus), 1–2, *2*, 71, 74
CPI (Consumer Prices Indexes), 378–379
Crash tests, 18–19
Creativity, 285–286, 584
Credit cards
  account numbers, 369
  cost of using, 186
  fraud, 27–28
  interest rates, 178, 186
  online security, 147
  understanding, 185–186

Cricket (Sport), 53–54
Critical path analysis, 588
Cross braces, 37
Cross hairs, 41, *41*
Cross products, 430–431
Cross-pollination, 443–445
Crowe, Russell, 228
Cryptography
  algorithms, 28
  for credit card security, 147
  development of, 116
  digital watermarks, 266–267
  discrete math for, 146
  domain and range of functions, 158
  for E-money, 160
  factoring and, 182
  group theory, 299
  inverse operations, 279, 280–281
  number theory and, 360, 362–363
  probability and, 424–426
  random number generators for, 75–76
  supercomputers for, 78
  word problems for, 585
  See also Encryption
Cryptosystem, 160
Cube root, **511–515**
Cubes
  architectural use of, 40
  Koch's, 350
  Rubik's, *243,* 243–244
  volume of, 439, 575
Cubic equations, 11, **438–440,** 439*t*
Cubic meters, 124, 129, 575
Cubing it (Power of three), 179
Cubit, 308
Cumulative frequency, 521
Currency exchange rates, 129, *195,* 195–196, *196,* 549, 586
Currency trading, 597
Curve fitting, 521–522
Curves
  area between, 91*f,* 92
  area under, 46, 83–85, 83*f,* 340–341
  calculus for, 80
  derivative of, 81–82
  description, 233
  French, 255
  maxima and minima, 85–86, 85*f*
  sine, 347, 494, 560–561, 560*f*
Cycling, 505–506
Cyclophosphamide, 324
Cylinders, 237

D

da Pisa, Leonardo, 13
da Vinci, Leonardo
  bicycle gears, 505
  Golden ratio, 443
  perspective, 391, *394,* 395
  proportion, 432, *433*
Dairy farms, 447
Dance, 266
Darwin, Charles, 174–175, 459
Darwinian evolution. *See* Evolution
Data
  authentication, 120
  fallacies, 256
Data compression, 118–119
  for compact discs, 118, 349
  for digital images, 264
  fractal, 202
  music, 349
Data encryption. *See* Encryption
Data encryption standard (DES), 362
Data mining, 28, 528
Data transmission, 29–30, 119–120
Dating
  absolute, 489
  carbon, 165, 298
  radioactive, 177
David (Michelangelo), *432*
Days, 120
De Bourcia, Louis de Branges, 212
De la Loubere's algorithm, 222
de Moivre, Abraham, 525
De Thiende (Viete), 139
de Vivie, Paul, 505–506
Death rates, 5, 513, 525
Debt, national, 5
Deceptive statistics, 381, 523–524, 581
Decibels, 297
Decimal fractions, 373
Decimals, 60, 61, **138–140**
  calculators and, 70
  conversion to percentage, 373
  fractions and, 204
  logarithms and, 295
  rounding off, 450
Decision making
  80/20 rule for, 370–371
  in football, 501–502
  game theory and, 228–229
  Prisoner's dilemma, 227, 596
  shopping, 181
  for zero-sum situations, 597
Declination, magnetic, 563
Decorating, 40
Decrease percentages, 375
Definite integrals, 83, 84*f*
Degrees
  of angles, 71, 557, 558
  of latitude and longitude, 103
Demographic profile, 141
Demographics, **141–143,** *142*
  rounding off, 451
  sampling for, 462
  school, 112–113
  See also Population growth
Denominators, 149
Dense-dose ACT, 323–325
Density, 124, 579
Department of Energy (U.S.), 214, 451
Dependent variables, 211, 249, 328
Depth of field, 400–401
Depth perspective, 263
Derivatives
  application, 86–91
  description, 81–83, 81*f,* 82*f,* 83*f,* 86
  higher-order, 83
DES (Data encryption standard), 362
DES (Digital encryption standard), 29
Descartes, René
  algebra, 13
  Cartesian coordinate system, 134
  exponents, 170
  geometry, 236
  graphing, 248, 257
  powers, 417
  root symbols, 512
Determinants, **303–306**
Determination, coefficient of, 404, 405
Deterministic models, 329
Deviation
  average, 520
  mean, 520
  standard, 318, 319–320, 520
Diabetes type II, 315
Diagnostic imaging and tests, 147, 264, 318–319, 475, 480
Diagrams, **404–415**
  body, 414
  development of, 407
  electric circuit, 414
  fishbone (Ishikawa), 406
  fundamental concepts, 404–407
  precedence, 588
  real-life applications, 407–415
  tree, 412–413
  for word problems, 583–584
Diatonic scale, 346
Dice, 216, 423–424
Dice Music (Mozart), 349
Die Coss (Rudolff), 512
Diet
  balanced, 436
  calorie counting, 3–4, *320,* 436, 534
  weight loss and, 534
Difference (Subtraction), 529
Diffie Hellman key agreement algorithms, 362
Digital cameras, 29, 154, 401
Digital cash. *See* E-money
Digital clocks, 139
Digital encryption standard (DES), 29
Digital images, 262–268
  blue screen, 134
  creating, 29, 263–264
  estimation in, 164–165
  forensic, 266
  fractal, 22, 202
  matrices and, 304–305
  processing, 403

quality of, 470–471
resolution of, 401, 470–471
Digital music, 348–349
Digital photographs, 134, 401
Digital signal algorithm (DSA), 363
Digital signatures, 363
Digital video disks (DVDs), 119
Dijkstra's algorithm, 588
Dimensions
of integers, 198
lesser, 232–234
matrices and, 303
measurement and, 308–309
See also Three-dimensions; Time
Diophantine approximation, 361
Diophantus, 12, 361, 417
Direct proportion, 431, 586–587
Dirichet, Johann Peter Gustav Lejeun, 157
Discrete mathematics, **144–148,** 474
Discrete objects, 144
Diseases
bacterial growth and, 481
causality of, 585
chronic, 322
demographics, 143
epidemiology, 314, 315–318, 322, 481
frequency, 315–316
game theory modeling, 230
genetic risk factors, 321–323
medical mathematics and, 314
outbreaks, 591
probability of, 208, 315–318
prognosis, 369
spread, 175
Disk rot, 88*f*
Dispersion, 520
Disraeli, Benjamin, 523
Distance
bar charts for, 111
braking, 408
camera lenses for, 402–403
conversions, 123, 128, 340, 545, 546*f*
definition, 307
depth of field and, 400–401
functions for, 211
illusion of, 391
mail delivery, 589, 589*f*
map scale and, 434
measurement of, 123, *123, 128,* 307, 308–309, 311
rounding off, 452
shortest, 308, 587–588, 587*f*
tables for, 545
three-dimensional measurement, 313
travel time and, 431
trip length, 443
Distribution
binomial, 317, 317*f,* 518
division and, 150–151
factoring and, 182
Gaussian, 518, 525
normal, 317–318, 318*f,* 518, 525
Pareto, 518
probability, 316–317
theoretical probability, 518
uniform, 518
Divergent sequences, 588
Divine proportion, 432, 436–437, 442
Division, **149–155,** 361, 486
DNA analysis. *See* Genetic sequence
DNA helix, 354
Doctor-patient ratio, 151
Doctrine of Chances (de Moivre), 525
Dog bites, 5
Domain, **156–158**
Domes, 237
Domestic animals, breeding, 483
Dominant inheritance, 444, 482
Doomsday Clock, 535
Double-axis line charts, 108–109
Doubling time, 173–174
Down payments, 187
Downhill ski racing, 339
Downloading, 118–119, 185
Doxorubicin, 324
Drag racing, 572
Drawings
perspective, 392, 392*f*
scale, 34–35
Dresher, Melvin, 596
Drinking water, 446
Drivers' licenses, 428
Driving
vs. air travel, 4, 370
fuzzy logic for, 301–302
See also Automobiles; Travel
Drogue parachutes, 17
Drowning, 5
Drugs
development, 147, 461
dosage, 47–48, 310–311, 434
manufacturing, 460–461
DSA (Digital signal algorithm), 363
DuBois formula, 48
Dummies, crash test, 18
Duomo, 389–390, 390*f*
Dürer, Albrect, 391
Durst, Seymour, 5
DVDs (Digital video disks), 119

E

E base, 297
Earned run average (ERA), 338–339, 498
Earnings, 66–67, 150–151
The Earth
age, 446
circumference, 105
curvature, 563
origin of the moon and, 447–448
surface area, *47*
volume, 162
Earth science, 489
Earthquakes, 289, *298,* 417, 471, 482–483
Eastman, George, 402
eBay, 230, 451, 531–532
Ecclesiastical modes, 345
Eclipse, 243
Ecological models, 330–331
Economics
game theory and, 228, 229
global, 515
Ecosystems, 199, 201
Edges (Graphs), 146
Education
income and, 193
school demographics, 112–113
student-teacher ratio, 445
tables for, 544–545
test score ranking, 589–590
word problems in, 584
See also Grades; Students
Edward I, 308
Efficiency, 584
Egyptians
algebraic development, 12
binary numbering systems, 336
calendars, 97–98, *98*
fractions, 205
geometry, 235
Golden ratio, 443
long division, 149
measurement systems, 235
multiplication, 336
pi, 442
powers, 416–417
proportion, 432
tables, 544
See also Pyramids, Egyptian
Eiffel Tower, 35–36, 38, *38*
80/20 rule, 370–371
Einstein, Albert
age of the universe, 160
spacetime and, 136
speed of light and, *514*
on time and space, 131–132, 236, 246
El Gamal algorithm, 363
Elapsed time, 3
Elections
ballot boxes, *7*
demographic analysis, 141–142
presidential, 141–142, 459
public opinion polls, 527–528
results, 383
sampling, 459
Electoral college, 209
Electric circuits, 414, 486
Electric current, 123, 124
Electric fields, 555
Electrical thermometers, 126

Electricity
magnetism and, 542
municipal supply, 589
prediction of demand, 93–95, 95*f*
solar panels for, 50
Electromagnetic radiation, 486–487
Electronic calculators. *See* Calculators
Electronic cash (E-money), 160
Electronic engineering, 489
Electronic musical instruments, 347–348
Electronic sound synthesizers, 213, 347–348
Electronic timing, 339
Electrons, 124
Elementary number theory, 360–361
The Elements (Euclid), 420–421
Elephants, legs, 179
Elevators, 359
ELF (Extremely low frequencies), 182
Ellipses, 233, 238, 346
Elliptic functions, **159–160**
E-mail, 341
E-money, 160
Empire State Building, 20
Employee pilferage, 153–154
Encryption, 28–29
128-bit, 425
data, 120–121
data encryption standard, 362
digital encryption standard, 29
number theory and, 362–363
probability and, 424–426
public key, 120, 147, 160, 280–281, 362–363
symmetric secret key, 120–121, 362
See also Cryptography
Encryption devices, 28–29
Endangered species, 587
Energy
consumption, 451
prediction of demand, 93–95, 95*f*
thermal, 125
Engine compression, 445, 579
Engineering
Domain and range of functions, 157–158
electronic, 489
factoring in, 182
measurement systems, 309
proportion in, 435
scientific notation, 487
vehicle weight and, 535
word problems, 585–586
English measurement system, 122, 123, 124, 472, 477
Eno, Brian, 349
Entropy, 274
Enumeration. *See* Counting
Environmental science, 488
Epidemics, 591
Epidemiology, 314, 315–318, 322, 481
Equations
algebraic, 10–12
cubic, 11, **438–440,** 439*t*
function graphs for, 255
functions as, 210–211, 211*t*
linear, 11, 287–293, 288*f,* 438–440, 439*t*
matrix, 288, 288*f*
Navier-Stokes, 212
nonlinear, 11
quadratic, **438–440,** 439*t*
quartic, **438–440,** 439*t*
reducing, 181–182
scientific use, 476
third-order, 438
Equilateral triangles, 237, 541, 558–559, 558*f*
Equilibrium, 229
Equivalent fractions, 204
ERA (Earned run average), 338–339, 498
Eratosthenes, 105, 420–421
Ergonomics, 42, 434–435
Error-correction codes, 269, 275–276, 349, 360
Errors
burst, 275
estimation of, 162
measurement, 309
medical, 5
RMS, 520
Estimation, **161–166**
Euclid, 236, 263, 420–421
Euclidean algorithms, 361
Euler, Leonhard, 156, 170–171
Euro, 196
Everest, 310, 565
Evolution
averages and, 57–58
game theory and, 228, 229–230
population growth and, 174–175
Evolution and the Theory of Games (Smith), 228
Execution grade, 504
Exemestane, 325, 326
Exercise, 3–4
Exercise equipment, 259
Exit polls, 142
Expenses, operating, 63
Exponential functions
description, 168–169, 169*f*
population growth, 171–173, 172*f,* 174, 200, 331, 338
for prediction of demand, 95
Exponentiation, 295
Exponents, **167–179**
development, 160–171
fundamental concepts, 167–169
growth rates and, 337
laws of, 169, 170*t*
real-life applications, 171–179
Expressions, variable, 10
Extreme programming (XP), 285
Extreme sports, 154
Extremely low frequencies (ELF), 182
Eyes, 281

F

Fabric design, 43
Face recognition, 264–266
Factoring, *10,* 11–12, **180–183,** 182
Fahrenheit, 125–126, 340, 357
Fahrenheit, Daniel Gabriel, 126
Failure prediction, 89–90, 427–428
Fallacies, 256
Family, average, 55–56
Family tree, 413
Fans, 41–42
Farad, 489
Farr, William, 322
Fast food restaurants, 4
Fast Fourier transform (FFT), 566
Fat, body, 578–579
FDA (Food and Drug Administration), 245
Feast of Herod (Donato), 391
Feature points, 291
Fermat's last theorem, 361–362
Fermat's principle, 89
Ferries, 563
FFT (Fast Fourier transform), 566
Fibonacci numbers, 353–354, 491, 492
Field, depth of, 400–401
Field theory, Abelian class, 361
Fighter aircraft guns, 41, *41*
Figure skating, 504
Files (Computer)
compression, 118–119
file-sorting programs, 425
forensic search of, 425
zipped, 119
Film
monomolecular, 162
photographic, 398–403
Films. *See* Motion pictures
Filtering, 48
FINA (International Amateur Swimming Federation), 387
Financial calculations, personal, **184–197**
Financial tables, 546–548
Finches, 58
Fingerprints, 22–23, 266
Finite sequences, 493
Finite-element models, 212–213
Fires, 5, 75
Fireworks, 5, 241
Fish populations, 164, 328–329
Fishbone diagrams (Ishikawa), 406
Fisher, R., 229, 525
Flash powder, 241
Flight insurance, 370
Flight mathematics, 479

Floating objects, 482
Flood, Merrill, 596
Floods, 250, 358–359
Floor numbering, 359
Flow charts, 411–412, 413*f*
Flow rate, 488
Flowers, 353–354, 539–540, 539*f*, 540*f*
Flowers, Tommy, 29
Floyd's algorithm, 588
Fluid mechanics, 281
Flying. *See* Air travel
Focal length, 399, 400*f*
FOIL method, 180–181
Food
  budgets, 194
  inspection technology, 266
  percentages, 374
  pH scale and, 418
  proportions, 431
  quality, 474
  rotting leftover, 173–174
  value meals, 4
  See also Diet
Food and Drug Administration (FDA), 245
Football
  coin toss, 500–501, 501*f*
  college championships, 14–15
  decision making in, 501–502
  fields, 386
  pass length, 507
  pattern recognition, 182
  quarterback rating index, 499
  salary caps, 507–508
  scores, 358
  statistics, 382, 499
Forces, 563–564
Forensic science
  digital images for, 266
  fingerprints, 22–23, 266
  forensic computing, 425
  scientific notation for, 488–489
Formulas, algebraic, 11
Fossil fuels, 49
Fossil samples, 460
401(k) plans, 191–192
Fourier, Jean Baptiste, 213, 347
Fourier transform spectroscopy, 566
Fourteenth Amendment, 209
Fourth dimension. *See* Time
The Fractal Geometry of Nature (Mandelbrot), 200
Fractals, *21*, 22, **198–202,** 199*f*
  in computer-generated music, 350–351
  in nature, 199, 200–201, 353, 355
  symmetry of, 541–542, 541*f*
Fractional voting rights, 209
Fractions, **203–209,** *204*
  continued, 206
  conversion to percentage, 373
  decimal, 138, 373
  definition, 203, 372
  division and, 149
  equivalent, 204
  improper, 203
  lowest-term, 204
  mixed, 204
  proper, 203
  vs. ratios, 442
  rounding off, 204–205
  rules for, 204, 205*t*
  in sports, 496
  unit, 203–204
Franklin, Benjamin, 162
F-ratio test, 522, 523
Fredn, William, 357
Free-throws, 500
Freezing point, 126, 356
French curves, 255
Frequencies (Sound), 241–242, 351
Frequency
  cumulative, 521
  of diseases, 315–316
  histograms, 410, 412*f*
  measurement of, 124
  of musical notes, 351
Friction, 563–564
F-stop, 400
Fuel consumption
  airplane, 16–17
  automobile, 341, 352, 383, 451
Fuel fluid mechanics, 281
Fuels. *See specific types of fuel*
Functional MRI, 264
Functions, **210–214,** 211*t*
  calculus and, 81, 81*f*, 82
  continuous, 255
  definition, 145, 156, 157
  domain and range of, 156–158
  elliptic, **159–160**
  graphing, 255
  inverse, 279, 279*f*
  scientific use of, 473–474
  sine, 560–561, 560*f*, 562
Fundamental theorem of calculus, 85
Fundraising, 19
Furlongs, 339
Furniture, 434
Fusion, 579–580
Future events, 6
Futures trading, 597
Fuzzy logic, 301–302

G

G note, 344
Galaxies, 55, 202, 486
Galley division, 149–150
Gallons, 340, 341, 452
Gallup polls, 459
Galois theory, 361
Galton, Francis, 373
Gamblers, compulsive, 367
Gambling and betting
  casino games, 219–220, 366–367, 425–426
  chance in, 424
  history of, 424
  Martingale system, 6
  odds, 366–367, 591
  probability myths, 425–426
  slot machines, 217–218, 219–220, 366–367
  sports, 223, 508–510
  zero-sum games, 596
Game math, **215–224**
  development of, 216–218
  fundamental concepts, 215–216
  potential applications, 223
  real-life applications, 218–223
Game theory, **225–231,** 332–335
Games
  board, 134, 220, *220*
  casino, 219–220, 366–367, 425–426
  computer, 217, 223, 396, 412
  musical, 349
  strictly competitive, 596
  zero-sum, 226, **595–597,** *596*
  See also Card games
Gantt graphs, 254, 254*f*, 413
Gases, greenhouse, 49
Gasoline
  consumption, 341, 352, 383, 451
  cost per trip, 443
  miles per gallon, 340, 341, 452
Gateway Arch, 238
Gauss, Karl Frederick, 162, 518, 525
Gaussian distribution, 518, 525
GCF (Greatest common factor), 181
Gears, bicycle, 505–506
Gender differences, 581–582
Gene expression profile, 327
General equilibrium, 229
General rubrics, 455–456
Generality, 329
Generational cohort, 143
Genetic sequence
  algorithms, 28, 30
  forensic, 489
  graphing, 258
  information theory and, 274–275
  potential applications, 327
  Punnet square for, 545, 546*f*
  use of, 493
Genetic traits, 321–323, 443–445, 482, 483
Genome, human, 327
  See also Genetic sequence
Geodesy, 564–565
Geographic coordinate systems, 133–134
Geographic information systems (GIS), 143
Geographic Poles, 563
Geologic time, 488

Geology, 591–592
Geometric forms
  architectural use, 40
  area of, 46*t,* 182, 340
  fractals as, 199
  sports and, 40–41
  symmetry of, 34
Geometric mean, 52, 519
Geometric number theory, 361
Geometric progression, 6
Geometry, **232–247,** *233, 238*
  analytic, 13
  development of, 235–236
  fundamental concepts, 232–235
  imaging and, 263
  potential applications, 245–247, 260
  real-life applications, 236–245
  visual, 35–36
Geostatistics, 525–526
GIF (Graphics interchange format), 165
Gigaelectron volts, 487
Girard, Albert, 13, 512
GIS (Geographic information systems), 143
Glaciers, 49
Glass thermometers, 126, 580–581
Glide-reflection symmetry, 537
Global economics, 515
Global positioning systems (GPS)
  automobile, 470
  continuously operating reference station, 565
  inertial guidance systems and, 93
  maps for, 105
  nautical navigation and, 74
  navigation and, 135–136
  potential applications, 566
  satellites, 239–241, 258
  surveying with, 50, 258
Global warming, 48–49, 257–258, 594
GMT (Greenwich Mean Time), 135, 547, 547*f*
Gold, 489
Goldbach, Christian, 421
Goldbach conjecture, 361, 421
Golden Gate Bridge, 76
Golden ratio, 38–40, 353, 354, 442, 492
Golden rectangle, 38–40, 41
Golf, 285, 358, 533–534
Golf clubs, 506–507
Gosset, William S., 525
GPS. *See* Global positioning systems
Grade of execution, 504
Grades
  average, 54
  grade point average, 139–140, 140*f,* 153
  percentages and, 374, 383
  scale of performance for, 454
  scoring rubrics for, 453–456
Gradians, 71
Grads, 558
Grant, Peter, 58
Grant, Rosemary, 58
Graph paper, 257
Graph theory, 147, 248, 588, 589
Graphic novels, 392
Graphic scale (Maps), 468
Graphical models, 329
Graphics
  computer, 396, 566, 572
  misleading, 581
  See also Digital images
Graphics interchange format (GIF), 165
Graphing, **248–261,** *249, 252*
  development of, 257
  discrete math for, 146
  Domain and range of functions, 158, 158*f*
  fallacies, 256
  functions, 255
  fundamental concepts, 248–257
  inequalities, 255
  real-life applications, 257–260
Graphing calculators, 69, *70,* 257
Graphs
  area, 252, 253*f,* 405
  bar, 249–251, 250*f,* 251*f,* 406, 410, 411*f*
  bubble, 257, 257*f*
  coordinate systems for, 248–249
  gantt, 254, 254*f,* 413
  legends and titles for, 404
  line, 251–252, 253*f,* 255, 408–410, 409*f*
  phylogenetic tree, 258
  picture, 254
  pie, 252–253, 253*f,* 406, 410, 410*f*
  radar, 253, 254*f*
  scatter, 404, 407–408, 408*f,* 409*f*
  scientific use of, 476
  stem, 250
  three-dimensional, 112, 407
  topology and, 554–555
  triangular, 407, 410–411
  x-y, 254–257, 254*f*
  x-y scatter, 109, 109*f*
  See also Charts; Diagrams; Plots
Graunt, John, 322, 525
Gravity, 16, 244, 313
Gray (Unit of measure), 124
Great Trigonometrical Survey of the Indian Sub-continent, 310, 564–565
Greatest common factor (GCF), 181
Greeks, 53, 420
Greenhouse gases, 49
Greenwich Mean Time (GMT), 135, 547, 547*f*
Gregorian calendar, 98, 99
Gregorian chants, 345
Grids, 37
Group theory, 542
Growth rates, 337–338
  See also Population growth
Guidance systems
  inertial, 91–93, 92*f,* 93*f*
  missile, 440
Guilloché patterns, 211–212
Guitar strings, 207, 242
Guitars, 242, 347
Gumball estimation, 163–164
Guns
  fighter aircraft, 41, *41*
  speed, 139

H

Half-life (Nuclear waste), 206, 213–214
Halley, Edmond, 525
HALO skydiving, 17
Hamming, Richard, 119–120
Hamming codes, 119–120
Hammond electric organs, 347
Handcuff puzzles, 555
Handicaps, 533–534
Hang time, average, *54*
Hard drive storage, 87–89, 88*f*
Harmonics, 345, 346
Harrison, John, 135
Hash functions, 363
Hawaiian-style guitars, 347
Hawk-Dove game, 229–230
Hazen's method, 521
Health. *See* Medical mathematics
Heat lamps, 20
Height
  arm span and, 408
  average, 368, 593
  bar graph of, 410, 411*f*
  box plot of, 407
  Mount Everest, 310
  to-weight ratio, 446
  and weight charts, 319–320, 548–549, 548*f,* 549*f*
Helix, 354
Henry IV, 279
Hercules, 338
Hertz (HZ), 124, 351
Heterozygous recessive inheritance, 321
Hexagons, 238, 239, *240*
High school demographics, 112–113
High-altitude, low-open (HALO) skydiving, 17
Higher-order derivatives, 83
Highways, toll, 533
Hillary, Edmund, 565
Hindu division methods, 149–150
Hindu-Arabic positional notation, 27
Hinsley, Harry, 29
Hipparchus, 134, 418, 561–562
Hippasus, 512, 513, 560
Hippocampus, 578
Hippocrates, 322
Histograms, 250–251, 404, 410, 412*f*
Historical conversions, 125
HIV (Human immunodeficiency virus), 322, 461

Hockey, ice, 496, 498, 502–503
Holistic rubrics, 455
Home accessories, 40
Home prices, 249–250, 250*f*, 251*f*
Homozygous dominant inheritance, 321
Homozygous recessive inheritance, 321
Honey bees, 239
Honeycomb, 239, *240*
Hooch, Pieter de, 391
Horizon, 389
Horse racing, 339, 425
Horses, 354
Hospital size, 440
Hot climate, 576
Hot numbers, 426
Hotelling, Harold, 525
Hour, miles per, 123, 128, 340, 452
Hours, 120
House edge (Casino gambling), 366, 367
Hubble Space Telescope, 55, 165–166
Human body
  body composition, 578–579
  body mass index, 214, 319
  body surface area, 48
  diagrams, 414
  proportions, 436–437
  reflection symmetry, 537, *538*, 542
  volume, 439
Human genome, 327
  See also Genetic sequence
Human immunodeficiency virus (HIV), 322, 461
Human motion tracking, 290–291
Human population growth, 173*f*, 174, 175
Huntington's disease, 321
Hurricane modeling, 201, 213
Huygens, 311
Hydra, 338
Hydrodynamics, 259
Hydrogen, 435, 486
Hyperinflation, 192
Hypersonic flight, 299
Hypotenuse, 559–560
Hypotheses, 522, 523, 583, 585
Hyuk, Catal, 467
HZ (Hertz), 124

I

Ice, polar and glacial, 48–49, 580
Ice hockey, 496, 498, 502–503
Identification bar codes, *23*
ILLIAC computer, 350
Illusions, optical, 394
Illustrations, 392
Imaging, **262–268**
  See also Digital images
Impeachment, 208–209
Improper fractions, 203
Inches, square, 45–46
Incidence, 316
Income
  average, 56, 152
  budgeting for, 188–189
  caps, 507
  education and, 193
  government salary tables, 548
  graphing, 250–251
  pay rate, 339
  salary caps, 507–508
Income taxes
  algorithms, 31–32
  calculating, 14
  deductions, 532
  deferred, 191–192
  negative numbers and, 356–357, 358
  payroll and, 65–66
  tables for, 548
  understanding, 189
Increase percentages, 375
Indefinite integrals, 84, 84*f*
Independent variables, 211, 249, 328
Indexes, 146, 590
Individual retirement accounts (IRA), 191–192
Inequalities
  graphing, 255
  linear, 290, 290*f*, 291–292
Inertial guidance systems, 91–93, 92*f*, 93*f*
Infants
  development, 343
  products for, 143
Infectious disease modeling, 230
Infinite sequences, 493
Infinite series, 22
Infinity, 150
Inflation, 177–178, 192
Information
  content, 272
  definition, 269–273
  meaning and, 273
  messages, 269–273, 271*f*
  quantum, 277
Information processing systems, 412
Information theory, **269–277**
  development of, 273
  error-correction codes and, 349
  fundamental concepts, 269–273
  real-life applications, 273–276
Inheritance
  dominant, 444, 482
  heterozygous, 321
  homozygous, 321
  Mendelian, 321, 443–445
  probability in, 483
  recessive, 444, 482
Injuries, sports, 573
Input, 328
Inspection technology, food, 266
Instructions, 414–415, 591
Insurance
  automobile, 57, 428
  averages and, 57
  flight, 370
  word problems for, 584–585
Integers
  definition, 145
  dimensions of, 198
  exponents and, 167–168
  greatest common factor, 181
  number theory and, 360–364
Integrals
  application, 91–96
  definite, 83, 84*f*
  description, 83–85, 84*f*, 86
  indefinite, 84*f*
Integration, 46, 577
Integrity, data, 120
Intelligence, artificial, 32, 230
Intelligence quotient (IQ), 555
Interactions and game theory, 225–231
Interest
  business math and, 67–68
  compound, 74–75, 170, 376, 591
  credit card, 178, 186
  exponential, 178
  exponents and, 177–178
  simple, 591
  tables of, 546
Interferometry, 309
Interior design, 394–395, 434, 594
Internal combustion engines, 445, 579
International Amateur Swimming Federation (FINA), 387
International Bureau of Weights and Measures, 129
International Organization for Standardization (ISO), 398
International Panel on Climate Changes, 580
International System of Units (SI), 122–130
International Ultraviolet Explorer satellite, 55
Internet
  data transmissions, 29–30
  IP address, 117–118
  online auctions, 230, 451, 531–532
  subnet mask, 118
  telephony, 119
  Voice over Internet protocol, 155
Internet Explorer, 425
Internet Protocol (IP) address, 117–118
Interval scale, 466, 468, 468*f*
Intervals, confidence, 522
Intuitive factoring, 182
Inverse, **278–283**
Inverse functions, 279, 279*f*
Inverse proportion, 431, 586–587

Inverse square law, 280
Investments
  description, 190–192
  growth rates, 337–338
  retirement, 192–194
  tables, 546–548
IP address, 117–118
IQ tests, 555
IRA (Individual retirement accounts), 191–192
Irish Lotto, 193–194
Irregular shapes, area of, 340
Ishikawa diagrams, 406–407
Isidore (Saint), 345
Islam
  calendars, 99
  trigonometry and, 562
ISO (International Organization for Standardization), 398
ISO number, 398–399, 401
Isosceles triangles, 559
Iteration, **284–286**
Ivan (Hurricane), 213

J

Jaguar population, counting, 147–148
James, Bill, 382
Jefferson, Thomas, 209
Jet fighters, 474, 478
Jewelry, 42–43, 577–578
Joachim, Georges, 562
Joint photographic experts group (JPG) format, 165, 264
Joists, 37
Jordan, Michael, 284
JPG (Joint photographic experts group) format, 165, 264
J-shaped growth, 330–331
Judges, sports, 504
Judgment sampling, 458, 461
Julian calendar, 98, 99, 451
Jumps, vertical, 534
Jurassic period, 488
Just in time manufacturing, 440

K

Kasparov, Gary, 217
Kelp bass, 328–329
Kelvin temperature scale, 123, 125, 126, 127, 357, 467
Kepler, Johannes, 346, 544
Keyboards, electronic, 348
al-Khwarizmi, Abu Abdullah Muhammad ibn Musa, 12–13, 26
Kilograms, 123, 124
Kilometers, 123, 128, 340
Kluge, General von, 332–335
Knight's Tour, 221
Knot theory, 355
Koch's cube, 350
Koch's curve, 199, 202
Kodak Cameras, 402
Kolmogorov, A. N., 525
Kruskal's algorithm, 589
Kurzweil, Ray, 57

L

Lacrosse, 496
Lake perimeters, 386–387
Lambdon, William, 310
Lambert, Johann Heinrich, 101, 407
Lamps, heat, 20
Land mines, 572–573
Land surveying. *See* Surveying
Landscaping, 386, 396
Language translation, 587
Laplace, Pierre Simon, 525
The Last Supper (Da Vinci), 395
Latitude, 103–104, 134
Lattice multiplication, 336
Laws of exponents, 169, 170*t*
The Laws of Thought (Boole), 145, 146
Leap years, 99
Leaves, 200, 353–354
Leech, John, 364
Lee-Kai-chen, 71–72
Leftover food, rotting, 173–174
Legs, 179, 354
Leibniz, Gottfried Wilhelm von
  calculus, 46, 80, 86, 87
  domain and range, 156
  functions, 210
  multiplication, 336
Leibniz notation, 86
Length, 123, 450
  See also Distance
Lens aperture, 400–403
Lenses, camera, 399, *399,* 400*f,* 402–403
Lesser dimensions, 232–234
Liberty Bell (Slot machine), 217
Licenses, drivers', 428
Lies, 381, 523–524
Life cycle, cicadas, 421–422
Life expectancy, 4, 57, 549, 550*f*
Light
  diagrams, 414
  speed of, 124, 312, 486–487, *514*
Light bulbs, 20–21
Light years, 486–487
Lightning, 193, 427
Likelihood. *See* Probability
Line charts, 107–109, 108*f,* 111
Line graphs, 251–252, 253*f,* 255, 408–410, 409*f*
Line of best fit, 404
Linear algebra, 571
Linear equations, 11, 287–293, 288*f,* 438–440, 439*t*
Linear inequalities, 290, 290*f,* 291–292
Linear mathematics, **287–293**
Linear programming, 291–292, 588
Linear scale, 465, 465*f,* 466*f*
Lines
  definition, 232
  graphing, 248
Linguistics, 31
Liquid volume, 340
Liters, 129, 340
Livestock production, 447
Loans
  automobile, 547
  for credit card purchases, 186
  interest on, 67–68
  student, 56–57
  tables for, 547–548
Lobachevsky, Nikolai Ivanovich, 157
Local analysis, 361
Lock (Odds), 366
Log tables, 296
Logarithmic scale, *298,* 465–466, 466*f,* 468*f,* 475
Logarithms, 72, **294–299,** 475
Logic, 144, **300–302,** 301*t*
Logical algorithms, 27
Logistic growth curve, 331
Long division, 149
Long multiplication, 336–337
Long odds, 366
Longitude, 103–104, 134, 135
The Lord of the Rings (Film), 394
Lorenz, Edward, 200
Lottery odds, 193–194, 207–208, 368–369, 591
Lowest-term fractions, 204
Luminous intensity, 123
Lunar calendars, 97–98, *98,* 99
Lunar cycles, 451
Lunar eclipse, 243
Lung volume, residual, 579
Lunisolar calendars, 98, 99

M

MAC address, 30
Machine parts, 579
Macros, 286
Madonna and Child with Saints (Masaccio), 391
Magic squares, 217, 221–223
Magic tricks, 428–429, 555–556
Magnetic declination, 563
Magnetic fields, 555
Magnetic poles, 563
Magnetic resonance imaging (MRI), 264

Magnetism and electricity, 542
Magnitude, 418, 569–570
Mail-in rebates, 532–533
Mailman routes, 589, 589*f*
Maine coastline, 198
Malaria, 321
Malls, 458
Malthus, Thomas Robert, 174–175
Mandelbrot, Benoit, 200, 350
Manhole covers, 236–237
Manufacturing, just in time, 440
Maps, **100–106**
  algorithms for, 30
  conformal, 159–160
  coordinate systems, 103–104, 131, 132–136
  development of, 104–105
  as diagrams, 413–414
  Domain and range of functions, 158
  Mercator projection, *101,* 101–102, 105, 134–135, 160
  ordinance survey, 136
  paper, 134–135
  projection of, 100–102
  scale, 100, 434, 443, 467–468
  street, 132
  topographic, 104, *104,* 136
  trigonometry and, 564–565
  weather, 414
  x-y graphs as, 256
Market analysis, 30, 143, 463
Marketing
  product samples for, 463–464
  viral, 341
Mars, 265, *265,* 462–463
Mars Climate Orbiter, 122
Martingale system, 6
Masaccio, 391
Mass, 123, 124, 129, 313
Matched sampling, 461
MathCad, 330
Mathematica, 330
Mathematical tables. *See* Tables
Mathematics
  business math, **62–68,** 182–183
  discrete, **144–148,** 474
  flight, 479
  game, **215–224**
  linear, **287–293**
  medical, **314–327,** 488
  photography, **398–403**
  rules, 495, 496–497
  scientific, 69, **473–483**
  skills, 544–545
  See also Architectural math; Sports math; specific topics
Math-rock music, 351–352
MatLab, 330
Matrices, **303–306,** 304*f,* 475
Matrix algebra, 146, 262, 303–304, 304*f,* 572
Matrix equations, 288, 288*f*
Maxima, 85–86, 85*f,* 518
Maximum lifespan, 57
Maximum likelihood estimation, 148
Maya calendars, 98
MBTF (Mean time between failures), 56, 427–428
MD5 algorithm, 425
Meals, value, 4
Mean, 51–53
  arithmetic, 51–52, 519
  deviation, 520
  geometric, 52, 519
  See also Average
Mean time between failures (MBTF), 56, 427–428
Meaning, 273
Mean-tone temperament, 346
Measles, Mumps, Rubella (MMR), 594
Measurement, **307–313**
  accuracy of, 309
  architectural, 35, 310
  calculators for, 75
  conversions, 122–130, 339–340
  decimal, 139
  Egypt and, 235
  English, 122, 123, 124, 472, 477
  historical conversions, 125
  International System of Units, 122–130
  precision of, 452
  United States and, 124–125
  See also Metric system
Measuring tools, 208
Mechanical calculators, 72
Mechanics, quantum, 277
Media
  news, 378–370, 429
  sports experts, 502–503
Median, 52
Medical errors, 5
Medical imaging, 264
Medical mathematics, **314–327,** *319*
  fundamental concepts, 315–318
  potential applications, 327
  real-life applications, 318–327
  scientific notation and, 488
Medical technology, 314
Medieval monks, 345
Memory cards, 154
Mendel, Gregor, 443–445, 459
Mendelian inheritance, 321, 443–445
Menelaus, 562
Mercator, Gerhardus, 135
Mercator projection, *101,* 101–102, 105, 134–135, 160
Mercury thermometers, 126, 580–581
Mesopotamians, 543–544
Mesozoic era, 488
Messages
  information, 269–271, 272
  unequally likely, 271–272
Metal density and volume, 577–578
Meters
  conversions, 128
  cubic, 124, 129, 575
  definition, 123
  history, 467
  per second, 123
  square, 124
Metric Conversion Act, 124–125, 308
Metric system
  conversions, 339–340, 477
  decimals and, 139
  description, 122–130, 308, 472
Mexican currency exchange rates, 195–196
Michelangelo, *432*
Microorganisms, 460
Microprocessors
  algorithms, 26–27
  in calculators, 69
  Moore's law of, 7
  Pascal, Blaise and, 3
Microscopes, optical, 403
Microwave ovens, 302
Miles per gallon, 341, 452
Miles per hour, 123, 128, 340, 452
Military cryptography, 78
Military models, 331–332, 333*f*
Military perimeters, 387–388
Miller, Frank, 393–394
Millibars, 469
Milliliters, 129
Millionaires, 192, 194
Milliradians (Mils), 558
Mills Novelty Company, 217
Minima, 85–86, 85*f,* 518
Minutes, 120, 187
Mir Space station, 428
Miracles, 426–427
Mirrors, 42–43, 395
Misleading graphics, 581
Misrepresentative sampling, 524
Missile guidance systems, 440
Mission-planning software, 258
Mississippi River, 488
Mixed fractions, 204
MMR (Measles, Mumps, Rubella), 594
Mobile phones. *See* Cellular telephones
Möbius strip, 555–556
Mode, 519
Modeling, 28, **328–334,** 353, 583
Models
  analytical, 329
  architectural, 433, 468
  automobile, 435
  deterministic, 329
  ecological, 330–331
  finite-element, 212–213
  graphical, 329
  military, 331–332, 333*f*
  of nature, 354–355
  population growth, 330–331, 332*f*
  predator-prey relationship, 422

Models (*contd.*)
simulation, 331–332
solar system, 471
three-dimensional, 78, 147, 468
Modulus, 361
Mole (Unit of measure), 123
Molecular cloning, 286
Molecular structure graphs, 147
Molecules
Avogadro's law of, 486
thickness, 162
Mollusk shells, 353, 354
Money
business math and, 62
counterfeit, 211–212
currency exchange rates, 129, *195,* 195–196, *196,* 549, 586
E-money, 160
geometric representation, 235
Monks, Medieval, 345
Monochord strings, 344
Monomolecular film, 162
Monophonic songs, 345
Monopoly (Game), 220
Monthly payments, 184–185, 187
Moon
calendars and, 97–98, *98*
lunar cycles, 451
orbit, 242–243
origin, 447–448
rock samples, 462–463
Moore, Gordon, 7
Moore's law, 7
Morbidity rates, 315
Morgenstern, Oskar, 226
Mortality rates, 5, 315, 525
Mortgage rates, 547–548
Moschopoulus, Emanuel, 217
Most valuable player, 381
Motifs, 43
Motion
second law of, 432
tracking, 290–291
Motion pictures
animation, 28, 134, 263, 391–393
perspective in, 393–394
Motorcycles, *54*
Mount Everest, 310, 565
Mountain height, 310
Mouse population growth, 338
Movement technique, 573
Movies. *See* Motion pictures
Moving average, 259
Mozart, Wolfgang Amadeus, 349, 432, 443
MRI (Magnetic resonance imaging), 264
Multiple axis line charts, 108
Multiple line graphs, 252, 253*f*
Multiple line payout machines, 219–220
Multiple radar graphs, 253
Multiplication, **335–342,** *336*
algebraic, 9–10
factoring and, 180–181
lattice, 336
long, 336–337
notations, 294–295, 335, 485–486
peasant method, 336
of vectors, 570–571
Multiplication table, *115,* **543–552**
development, 72, 335–336
log tables and, 296
student use, 544–545, 545*f*
Multiplicative inverse, 278, 281
Multiplying machines, 336–337
Municipal electric service, 589
Murals, 396
Music, **343–352**
buying, 184–185, *185*
compression of, 349
computer-generated, 349–350
development of, 343–347
digital, 348–349
downloading, 185
electronic, 213, 347–348
fractions and rhythm, 206–207
fundamental concepts, 343
intensity of, 297
linear reproduction, 292–293
math-rock, 351–352
notation systems, 345–346, *350*
pitch, 445
potential applications, 352
randomness in, 349, 351
real-life applications, 347–352
rhythm, 445
rock, 297, 351–352
stereo sound, 282–283
Western, 351
Music for Airports (Eno), 349
Musical games, 349
Musical instruments
acoustic, 242
electronic, 347–348
proportion in, 435
scale theory and, 467
Musical iteration, 285, *285*
Musical notes, 207, 344, 351
Musical scale, 207, 344, 346–347, 351, 471
Musical scores, 345–346, *350*
Musikalisches Wurfelspiel (Mozart), 349
Muslims. *See* Islam

N

Names, drawing, 424
Nanometers, 490
Nanotechnology, 418, 490
Napier, John, 72, *115,* 296
Napier's bones, *115*
NASA. *See* National Aeronautic and Space Administration
Nash, John Forbes, 228, 229
Nash equilibrium, 229
National Aeronautic and Space Administration (NASA)
Apollo mission, 535
Challenger Space Shuttle, 56, 427
Columbia Space Shuttle, 56
error-correction codes, 276, 364
failure probability and, 53, 427
Mars Climate Orbiter, 122
rocket launches, 480–481
space distance measurement by, 312
Space Shuttle, 56, 305, 427, 479, 481
Viking 1 and *Viking 2* missions, 265, *265*
Voyager spacecraft, 276
weather patterns and, 477
See also Space travel
National Collegiate Athletic Association (NCAA), 14–15
National debt, 5
National Institute of Standards and Technology (NIST), 29
National Safety Council, 5
National Security Agency, 147
Natural and Political Observations Upon the Bills of Mortality (Graunt), 525
Natural perspective. *See* Perspective
Natural resource maps, 105
Nature, **353–355**
fractals in, 199, 200–201, 353, 355
Golden ratio in, 442–443
Nautical navigation, 73–74, 135
Navier, Claude-Louis, 212
Navier-Stokes equations, 212
Navigation
astronomical charts for, 297
azimuths for, 562–563
bearings for, 562–563, 592
clocks and, 135
coordinate systems for, 131, 134, 135
global positioning systems and, 135–136
measurement systems for, 311–312
nautical, 73–74, 135
Pythagorean theorem and, 514
for shortest distance, 587–588, 587*f*
in space, 93
Nazi encryption codes, 29
NCAA (National Collegiate Athletic Association), 14–15
Negative numbers, **356–359,** *357*
history, 13, 279–280, 356–357
square root of, 512
Negative predictive value, 318–319
Negatives, photographic, 281
Networks, computer, 259–260, 555
Neumann, John von, 226, 424, 596
New Year's resolutions, 534
Newman, M. H. A., 29

News media
probabilities, 429
statistics, 378–379
Newton, Isaac
calculus and, 46, 80, 86, 87
domain and range, 157
gravity and, 313
interferometry and, 309
polar coordinate system and, 134
relativity and, 246
second law of motion, 432
square root and, 512
trigonometry and, 562
Newton's method, 86
NIST (National Institute of Standards and Technology), 29
Nitrogen, 435
Nodes graph, 146
Noise
brown, 350
reducing, 182
Noisy channels, 274
Nominal scale, 467
Non-integer exponents, 168
Nonlinear design, 281
Nonlinear equations, 11
Nonliving systems models, 201–202
Non-probability sampling, 458–459
Norgay, Tenzing, 565
Normal distribution, 317–318, 318*f*, 518, 525
North Geographic Pole, 563
North Magnetic Pole, 563
Notation systems
Hindu-Arabic positional, 27
logarithms and, 294–295
musical, 345–346, *350*
scientific, 171, 452, **484–490**
set, 492–493
subtraction, 531
Notes (Music), 207, 344, 351
Novels, graphic, 392
N-tiles, 521
Nuclear power plants, 311, 370
Nuclear waste
half-life, 206, 213–214
radioactive decay, 175–177, *176*, 176*f*
Null hypothesis, 522, 523
Number theory, 145, **360–364**
algebraic, 361
analytic, 361
continued fractions in, 206
cryptography and, 147
elementary, 360–361
game math and, 215
geometric, 361
Numbering systems
base 5, 59–60, 60*f*
base 8, 60, 61
base 20, 60
base 60, 60, 557–558
base e, 297
See also Binary numbering system; Decimals
Numbers
account, 369
composite, 180
Fibonacci, 353–354, 491, 492
hot, 426
negative, 13, 279–280, **356–359,** *357*, 512
perfect, 361
rounding off, 449–452
Social Security, 190
whole, 449–450
See also Prime numbers
Numerators, 149

O

Object oriented programming, 330
Objects
continuous, 144
discrete, 144
floating, 482
ordering, 493
packing, 590
See also Topology
O'Brien-Fleming stopping boundary, 326
Obtuse angles, 558
Obtuse triangles, 559
Ocean levels, 580
Octaves, 344, 346
Odds, **365–371**
betting and, 223
definition, 216
Domain and range of functions, 157
game math and, 215–216
long, 366
percentages and, 373
sports betting, 508–509
See also Probability
OHAHOA, 564
Ohms, 489
Oil
consumption, 451
exploration, 258, 366
on water, 162
Oil change, 154–155
Omnibus Trade and Competitiveness Act, 125
On-base percentage, 498
1% method, 374
128-bit encryption, 425
O'Neill, Tip, 53
Online auctions, 230, 451, 531–532
Online commerce, 4–5, 147
Online security, for credit cards, 147
Operating expenses, 63
Opinion polls, 379, 527–528, 593
Optical disks. *See* Compact disks
Optical equipment, 298
Optical illusions, 394
Optical microscopes, 403
Optics, mathematical, 264
Optimal solutions, 584
Optimization, 329
Orbit (Planetary), 162, 346
Ordering objects, 493
Ordinal scale, 467
Ordinance survey maps, 136
Ordinates, 254, 256
Organizational charts, 413
Organs, Hammond electric, 347
Oscilloscope, 347
Out-of-focus area, 403
Output, 328
Ovals, 233
Oven temperature, 127, *127*, 302
Overtime pay, 208
Overtones, 347
Over/under bet, 509
Overweight persons, 214
Owl, spotted, 305
Oxford English Dictionary, 485

P

Paciolus, Lucas, 13
Packing, 590
Paclitaxel, 324
Paintings
digital images of, 267
perspective, 391–392, *394*, 395
proportions, 21–22
scale, 467
Paints, color mixing, 22
Paper, graph, 257
Par scores, 358
Parachutes, 17–18
Paradox, 588
Parameters, 328
Pareto, Vilfredo, 518
Pareto distribution, 518
Parity code, 119
Parlay, 509–510
Parthenon, 38–40, *39*, 432, 433
Parts per million (PPM), 151–152
Pascal, Blaise
adding machines and, 2–3
calculators and, 72
probability and, 424
triangles and, 500–501, 501*f*
Pascal (Unit of measure), 124
Pascaline, 72
Pascal's triangle, 500–501, 501*f*
Passwords, 145
Pattern recognition, 182
Patterns
identification of, 181
in nature, 355
repeating, 284–286
voting, 383

Pay
overtime, 208
rate, 339
See also Income
Payoff, 226
Payroll, 65–66, 66*f*
PCAHs (Polychlorinated aromatic hydrocarbons), 578
PCR (Polymerase chain reaction), 489
PDOP (Positional dilution of precision), 258
Pea plants, 443–445
Pearson, Karl, 525
Pearson correlation coefficient, 521
Peasant method (Multiplication), 336
Pendants, 42
Pendulum clocks, 311
The Pentagon (Washington, D.C.), 40, *234*
Pentagons, architectural, 40
Percentages, **372–384, 375**
completion, 499
development, 373–374
fractions and, 204–205
fundamental concepts, 372–373
odds and, 366
on-base, 498
potential applications, 383–384
in sports, 496
winning, 154
wording of, 586
Percentiles, 372
Perfect numbers, 361
Perfect square, 181
Performance
athletic, 152, 154, 445–446, 495, 497–498
scale of, 454
Perimeter, **385–388,** *387*
Perimeter barriers, 386, 387–388
Perimeter detection systems, robotic, 388
Periodic tables, 545, 546*f*
Periscopes, 41
Personal computers. *See* Computers
Personal financial calculations, **184–197**
Perspective, **389–397**
atmospheric, 391
depth, 263
development, 389–391
fundamental concepts, 389
real-life applications, 391–396
Peso, 195–196
Petals, flower, 539–540, 539*f*, 540*f*
PGP (Pretty Good Privacy), 281
pH scale, 418
Phi, 354, 436
Phones. *See* Cellular telephones
Photographic film, 398–403
Photographic negatives, 281
Photographs
black and white, 281
digital, 134, 401
out-of-focus, 403
sports, 402–403
wildlife, 402–403
See also Digital images
Photography math, **398–403,** *402*
Photomicrography, 403
Photovoltaic panels, 50
Phyllotaxis, 354
Phylogenetic tree graphs, 258
Physical fitness equipment, 259
Physical modeling synthesis, 213
Physics, 157, 274
Phytoplankton, 354–355
Pi
area of a circle and, 340
calculators and, 71
Egyptians and, 235
history of, 442
rounding off, 450
Picture graphs, 254
Pie charts, 110–112, 111*f*, 113, 113*f*
Pie graphs, 252–253, 253*f*, 406, 410, 410*f*
Pilferage, 153–154
Pitch (Musical), 445
Pitchers earned run average, 338–339, 498
Pitches, speed of, 139, 505
Pixels, 117, 164–165, 262, 264
Pixels per inch (PPI), 401
Planck constant, 124, 487
Plane angles, 123, 557–558
Plane triangles, 558–559
Planes (Air). *See* Aircraft
Planes (Geometry), 233
Planets
formation, 242
orbit, 162, 346
shadows of, 478
tables of, 544
Plants, 460
Players, most valuable, 381
Plots, **404–415**
box, 404, 407, 408*f*
development, 407
fundamental concepts, 404–407
real-life applications, 407–415
scatter, 408*f*
stem and leaf, 404, 407
Plutonium 238, 176–177, 176*f*, 206, 213–214
Poincaré, Jules Henri, 199
Point guards, 445–446
Point spread, 509
Points, 232, 248, 291
Poker, 5, 219, 366, 596
Polar bears, 576
Polar charts, 406–407
Polar coordinate system, 132, 133, 133*f*, 136–137
Polar ice, 48–49, 580
Pollination, 443–445
Pollock, Jackson, 200, *200*
Polls, public opinion, 379, 459, 527–528, 593
Pollution, 462
Polo, Marco, 216
Polychlorinated aromatic hydrocarbons (PCAHs), 578
Polygons, 233, 385–386, 417
Polymerase chain reaction (PCR), 489
Polynomials, 512
Polyphonic songs, 345
Polyrhythm, 351
Pools, swimming, 152, 386–387, 446, 581
Population dynamics, 22, 305, 330–331
Population growth
bacterial, 173–174, 326–327, 326*f*, 481
evolution and, 174–175
exponential, 171–173, 172*f*, 174, 200, 331, 338
Fibonacci numbers and, 353
human, 173*f*, 174, 175
logistic growth curve, 331
modeling, 330–331, 332*f*
of rabbits, 171–173, 172*f*, 200, 338, 353, 408
S-shaped, 331
Tasmanian sheep, 331
Populations
decrease in, 587
demographics, 141–143, *142*, 451, 462
modeling, 330–331
sampling, 164, 516–517
school, 112–113
Positional dilution of precision (PDOP), 258
Positional notation, Hindu-Arabic, 27
Positive predictive vale, 318–319
Pothole covers, 236–237
Pound, British, 196
Powder, black/flash, 241
Power plants, 94–95, 370
Power of three (Cubing), 179
PowerBall, 193–194
Powers, 296, **416–419**
PPI (Pixels per inch), 401
PPM (Parts per million), 151–152
Precedence diagrams, 588
Precision, 452
Predator-prey relationship, 422
Prediction, 6, 329, 494
Predictive value, 318–319
Premiums, 57
Presidential elections, 141–142, 459
Pressure measurement, 124
Pretty Good Privacy (PGP), 281
Prevalence, 316
Price tags, 4
Pricing
discounts and markups, 376–377
percentages, 374
proportion and, 431
by volume, 577–578

Prime factorization, 361
Prime numbers, **420–422,** 421*t*
definition, 145, 420
elementary number theory and, 360–361
factoring and, 180
highest, 421, *421*
Riemann hypothesis, 212
The Principles of Algebra (Fredn), 357
Prisoner's dilemma, 227, 596
Probability, **423–429,** *426*
addition and, 5
betting and, 223
card games and, 219
combinatorics for, 145
of disease, 208, 314, 315–318
distribution, 316–317
fundamental concepts, 215–216
genetic traits and, 321, 482, 483
logic and, 301
lottery winning and, 207–208
myths, 425–426
odds and, 365–371
relative frequency, 517–518
sampling, 457–458
scientific use of, 474
statistics and, 517–518
wildlife populations and, 147–148
Problem solving iteration, 284
Problems. *See* Word problems
Profits
business math and, 66
calculating, 531–532
derivatives for, 86–87
maximizing, 588
Prognosis, 369
Programming (Computer)
extreme, 285
linear, 291–292, 588
for modeling, 330
object oriented, 330
word problems for, 584
Progressive machines, 220
Project management, 285
Projection, Mercator, 100–102, *101,* 105, 134–135, 160
Prokofiev, Sergei, 349
Promotion (Game), 216
Proof, 583
Proper fractions, 203
Property line perimeters, 386, *387*
See also Surveying
Proportion, **430–437**
architectural, 34, *34,* 38, 432–433
definition, 372, 430
direct, 431, 586–587
divine, 432, 436–437, 442
epidemiology and, 315
inverse, 431, 586–587
percentages and, 373
wheels and, 43
Proportionality constant, 431
Propositions, 300, 301*t*
Proteins, 490
Proterozoic eon, 488
Protons, 124
Ptolemy, 562
Public key encryption
description, 120
for E-money, 160
inverse operations, 280–281
number theory and, 362–363
for online shopping, 147
Public opinion polls, 379, 527–528, 593
Punnet square, 545, 546*f*
Purchases
bulk, 450
rounding off, 451–452
savings per, 341
Puzzles
handcuff, 555
magic squares, 221–223
math, 223
Rubik's cube, *243,* 243–244
topographical, 555
Pyramids, Egyptian
area of, 46
geometry of, 35, 235, 237
proportion of, 432
ratio and form, 38
word problems for, 584
Pythagoras
architecture and, 36
geometry and, 236
music and, 343–345
proportion and, 431–432
root symbols and, 512
Pythagorean theorem
bridge span and, 478
camera lenses and, 399
carpentry and, 480, 480*f*
description, 512, 513, *513*
development of, 236
Navigation and, 514
number theory and, 361
right triangles and, 559–560
scientific use of, 474–475, 475*f*
space exploration and, 478
three-dimensions and, 313
trigonometry and, 474–475, 475*f*
weather balloons and, 478

Q

Quadratic equations, **438–440,** 439*t*
Quadrivium, 345
Qualitative statistics, 141, 455, 474
Quality control, 309, 526–527, 528, 592–593
Quantiles, 521
Quantitative statistics, 141, 455
Quantum computers, 276–277
Quantum cosmology, 487
Quantum information, 277
Quantum mechanics, 277
Quarterback rating index, 499
Quartic equations, **438–440,** 439*t*
Quartz clocks, 311
Qubits, 277
Quota sampling, 458, 459
Quotients, 442

R

Rabbit population growth, 171–173, 200, 338, 353, 408
Racing
drag, 572
horse, 339, 425
measurement of, 311
rally, 586
ski, 339
track and field events, 3, 358, 496, 534
Radar
graphs, 253, 254*f*
polar coordinate system and, 136–137
stealth technology and, 244–245
Radians, 71, 123, 558, 564
Radiation
absorbed dose, 124
electromagnetic, 486–487
exposure, 370
shielding, 299
Radiation therapy, 298–299, 325
Radiators, automobile, 49
Radio antennas, 182, 202
Radio technology, 542
Radioactive dating, 177
Radioactive decay, 175–177, *176,* 176*f*
Radioactive waste, 206, 213–214
Rafters, 237–238
Railway, 590, 590*f*
Rain
acid, 418
on aircraft, 474, 479
graphing, 249
measurement, 77
Rakine scale, 126
Rally racing, 586
Rand Corporation, 227, 596
Random number generators, 75–76, 517
Random sampling, 457–458, 460, 462, 517
Randomness, musical, 349, 351
Range, **156–158,** 518
Rankine scale, 125, 127
Ranking test scores, 589–590
Rates
base, 374–375
definition, 374
epidemiology and, 315
of increase/decrease, 375

Rates of change. *See* Derivatives
Ratings percentage index (RPI), 503–504
Rationing, 151
Ratios, **441–448,** *444*
architectural, 33–34, 38, *38*
cross products and, 430–431
epidemiology and, 315
examples of, 151–152
golden, 38–40, 353, 354, 442, 492
percentages and, 373
proportion and, 430–431
scale, 466–467
in sports, 496
trigonometry and, 562
wheels and, 43
Reaction time, 505
Réaumur scale, 125, 128
Rebates, 377–378, 532–533
Rebounds, 500
Recessive inheritance, 444, 482
Recipes, 206, 446, 591
Reciprocity, photographic, 401
Recorde, Robert, 13
Recordings, analog, 348
Rectangles
architectural use of, 238
area of, 45, 340
golden, 38–40, 41
perimeter of, 385
Reducing equations, 181–182
Redundancy, triple, 275
Reed-Solomon codes, 276
Refinements, 583
Reflection symmetry, 38, 537–542, 538*f*
Regression, 405, 406, 521–522
Relationships
functions and, 210
scientific, 474, 476–477
Relative frequency probability, 517–518
Relativity and time, 131–132, 246
Religious charts, 297
Remote vehicles, 82
Rendering, 134
Rentals, 384
Repeating patterns, 284–286
Repetition algorithms, 119
Representative fraction scale, 468
Representatives, Congressional, 209
Research sampling, 316
Residual lung volume, 579
Resolution
computer monitor, 117
digital image, 401, 470–471
Resolutions, New Year's, 534
Restaurants
budgeting for, 194
calculating tips, 194–195, 375–376
fast food, 4
Restitution, coefficient of, 504–505, 506–507
Retail sales analysis. *See* Sales analysis
Retirement accounts, 190, 191–194
Retreat, 229–230
Revenue generation, 63
Reverberation, 242
Revere, Paul, 271–272, *273,* 275
Revolutions per minute (RPM), 445
Rewards, 6, 191
Rey, Charles, 217
RGB scheme, 262
Rhind pyrus, 417
Rhythm, 206–207, 351, 445
Richter scale, *298,* 417, 471, 482–483
Riddles, 593
Riemann, Georg Friedrick Bernhad, 212
Riemann hypothesis, 212
Right angles, 558
Right triangles, 559–560, 559*f*
Risk
aversion, 597
calculating, 157
disease, 315–318
equals reward principle, 191
investment, 190–191
Rituparna, 53
Rivers, 488, 563
Riveste, Ronald, 28, 362
RMS errors, 520
Roads, toll, 533
Robotic perimeter detection systems, 388
Robotic surgery, 245
Rock music, 297, 351–352
Rock samples, 462–463
Rockets, 89, 274, 480–481
Roman numerals, 6–7
Romans, fractions and, 205
Roosevelt, Franklin, 190
Root mean square, 520
Ropes, 537, 539, 539*f*
Rotation, 34, 566
Rotational symmetry, 38, 537–542, 539*f,* 540*f*
Roth, Al, 230
Roth Individual Retirement Accounts, 191–192
Rotting leftover food, 173–174
Roulette, 219, 366, *367*
Round objects. *See* Circles; Spheres
Rounding, 204–205, **449–452**
Rovers, unmanned, 388
RPI (Ratings percentage index), 503–504
RPM (Revolutions per minute), 445
RSA algorithm, 28, 362, 363
Rubik, Erno, 243
Rubik's cube, *243,* 243–244
Ruble, 196
Rubric, **453–456,** *454, 455*
Rudolff, Christoff, 512
Rudolphine tables, 544
Rule of similar triangles, 182
Rules math, 495, 496–497
Run charts, 409–410
Run length compression, 119
Running tracks, 309
Runs scored, 498
Russian multiplication, 336
Ruth, George (Babe), 517
Rutherford, Ernest, 446

S

Sabermetics, 382, 498
Sabine formula, 242
Salamis Tablet, 71
Salaries. *See* Income; Pay
Salary caps, 507
Sales analysis
algorithms, 30
percentages for, 375
price discounts and markups, 376–377
weighted averages, 54–55
Sales tax, 377–378
Salesperson, traveling, 589
Sampling, **457–464,** *459,* 472
misrepresentative, 524
populations, 516–517
research, 316
San Francisco earthquake, 417
Sargon of Akkad, 105
SAT (Scholastic Aptitude Test), 382–383
Satellites
antennas, 536
error-correction codes, 364
GPS, 239–241, 258
International Ultraviolet Explorer, 55
spacetime and, 136
Savings accounts, 74–75
Savings per purchase, 341
Scale, **465–472,** *470*
bar, 468
grading, 453
graphing, 248
interval, 466, 468, 468*f*
linear, 465, 465*f,* 466*f*
logarithmic, *298,* 465–466, 466*f,* 468*f,* 475
map, 100, 434, 443, 467–468
musical, 207, 344, 346–347, 351, 471
nominal, 467
ordinal, 467
of performance, 454
ratio, 466–467
representative fraction, 468
by telescopes, 43
Scale drawings, 34–35
Scalene triangles, 559, 559*f*
Scanners, fingerprint, 22–23
Scatter graphs, 404, 407–408, 408*f,* 409*f*
Scatter plots, 408*f*
Schedules, 549, 550, *551*
Scholastic Aptitude Test (SAT), 382–383

Schools. *See* Education; Students
Scientific math, **473–483,** 475*f*
calculators for, 69
development, 476
fundamental concepts, 473–476
real-life applications, 476–483
Scientific notation, 171, 452, **484–490**
Scientific relationships, 474, 476–477
Scientific visualization, 258, 260, 260*f*
Scores
musical, 345–346, *350*
par, 358
sports, 3
test, 524, 589–590
Scoring rubrics, 453–456
Scratch division, 149–150
Screening tests, 318–319, 320*t*
Sculpture and proportion, 433–434
Sea levels, 580
Sea waves, seismic, *298,* 471, 591–592
Search engines, 146, 302
Seasons, 98
Seatbelts, 18
Seats, car, 42
Seawater, 489
Secant, 564
Second law of motion, 432
Seconds, 120
Secret writing. *See* Cryptography
Security, for credit cards, 147
Seedheads, 354
Seeding algorithms, 31, *31*
Seismic sea waves, *298,* 471, 591–592
Self-check-out, 368
Semicircular arches, 237
Semiconductors, 292
Sensitivity, 318–319, 320*t*
Sensors
charged-coupled device, 401
complementary metal oxide semiconductor, 401
crash test, 18
perimeter detection, 388
Sequences, sets and series, 144, **491–494,** 493*f,* 588
Sex ratio, 447
Shamir, Adi, 28, 362
Shannon, Claude
Boolean algebra and, 146
computers and, 116, 119
error-correction codes and, 349
information theory and, 269, 273, 274
Shape of objects. *See* Topology
Shapley, L. S., 228
Shark attacks, 193
Shatranj, 216
Shaturanga, 216
Sheep, 331
Shells, 353, 354, *354*
Ships, 482
Shoeprints, 266
Shooting percentages, basketball, 374
Shopping decision making, 181
Shubik, M., 228
Shutter speed, 399, 401
SI (International System of Units), 122–130
Sickle-cell anemia, 321
Sierpinski triangle, 541–542, 541*f*
Sieve of Eratosthenes, 420–421
Sigmoid growth curve, 331
Signatures, digital, 363
Significance, 522
Signs, street, 414
Silence, 345–346
Silo, 237
Similar triangles, rule of, 182
Similarity fractals, 199, 199*f*
Simple interest, 591
Simpson, Joanne, 77–78
Simulation models, 331–332
Sin City (Miller), 393–394
Sine curves, 347, 494, 560–561, 560*f*
Sine function, 560–561, 560*f,* 562
Sinusoidal equal area projections, 102, *102*
Sinusoids, 213
Sirius, 97–98
Six Sigma, 526–527
Skating, figure, 504
Ski racing, 339
Skin area, 48
SKIPJACK algorithm, 29
Skydiving, 17–18
Skyscrapers, 19–20
Slaves, 209
Sleep management, 534
Slide guitars, 347
Slide rules, 72
Sliding symmetry, 34
Slot machines, 217–218, 219–220, 366–367
Smith, John Maynard, 228
Smoking, 4, 457–458
Sniping, 230
Snooker, 279
Snow, John, 322
Snowball sampling, 459
Soccer, 506
Social Security number, 190
Social Security system, 190, 193
Soft drinks, 524
Software
architectural, 470
calculator, 72–73
for charts, 112
creative design, 584
development, 166
domain and range of functions, 158
face recognition, 264–266
graphing, 260
internal logical organization, 412
mission-planning, 258
modeling, 330
probability and, 429
spreadsheet, 257, 260, 549–550
trial, 285
visualization, 258, 260, 260*f*
word problems in, 584
SOHCAHTOA, 564
Soil tests, 410–411, 459–460, 472
Solar calendars, 98
Solar eclipse, 243
Solar panels, 50
Solar systems, 242–243, 471
Solid angles, 123
Solid objects, 46, 417
Solutions, optimal, 584
Songs, 185, 345, 445
See also Music
Sonoluminescence, 579–580
Sorites paradox, 588
Sorting algorithms, 590
Sound
amplification, 347–348
analysis, 347
compression, 349
frequencies, 241–242, 351
intensity, 297
reducing, 182
sine waves, 494
spacial geometry, 241–242
stereo, 282–283
synthesis, 213
Space
architectural, 37
sound manipulation with, 241–242
time and, 236
Space oblique Mercator projection, 102
Space Shuttle
Challenger, 45, 427
Columbia, 45
computer steering, 305
failure prediction for, 56, 427
launches, 481
See also National Aeronautic and Space Administration
Space Station Mir, 428
Space travel
antennas and, 535–536
distance of, 311, 312
navigational calculations for, 93
private, 24, *24*
Pythagorean theorem and, 478
timekeeping and, 312
unmanned vehicles, 388
Spacecraft
error-correction codes for, 276
private, 24, *24*
radiation shielding for, 299
scientific visualization, 260
steering, 305
vehicle weight, 535
vibration, 479

SpaceShipOne, 24, *24*
Space-time continuum, 246–247
Spam, 341
Spatial correlation, 525–526
Specificity, 318–319, 320*t,* 329
Spectra, 90–91
Spectroscopy, 566
Speed
derivatives for, 83
of gravity, 313
of light, 124, 312, 486–487, *514*
measurement, 123, 151, 307, 311
of photographic film, 398–399
of pitches, 139, 505
revolutions per minute and, 445
vectors, 233–234
See also Velocity
Speed guns, 139
Speed limits, 154
Speed-distance-time calculators, 74
Spheres
chords and, 562
harmony of, 346
measurement, 75
projection of, 101–102
volume, 576
Spherical triangles, 559
Spider (Radar) graphs, 253, 254*f*
Spirals, 354
Splines, 255
Sports injuries, 573
Sports judges, 504
Sports math, **495–510,** *497*
athletic performance, 152, 154, 445–446, 495, 497–498
attendance statistics, 409
betting, 508–510
championship standings, 31, *31,* 382
development, 496–497
field perimeters, 386
fundamental concepts, 495–496
geometric forms, 40–41
judging and, 504
percentages in, 374
photographs, 402–403
probability in, 426–427
ratings percentage index, 503–504
real-life applications, 497–510
rules, 495, 496–497
salary caps, 507–508
scoring systems, 3
statistics, 380–382, 498
surveying and, 514–515
timing systems, 139
tournament standings, 590
video analysis, 263
See also specific sports
Sports media experts, 502–503
Spotted owl, 305
Spreadsheet software, 257, 260, 549–550
Square centimeters, 45–46
Square inches, 45–46
Square meters, 124
Square root, **511–515**
Squares
area of, 46, 340, 439
magic, 217, 221–223
perfect, 181
perimeter of, 385
reflection symmetry, 540, 540*f,* 541
rotational symmetry, 541
Squaring it, 179
S-shaped growth, 331
St. Louis Gateway Arch, 238
Stacked bar graphs, 251
Stacked column charts, 110, 110*f*
Stadiums, 433
Staffing levels, 589
Stained glass, 42–43
Standard deviation, 318, 319–320, 520
Star atlas, 55
Star (Radar) graphs, 253, 254*f*
Stars (Astronomy), 242, 418
Star-shaped fireworks, 241
State Plane Coordinate System, 103–104
Statistics, **516–528,** *518, 523*
deceptive, 381, 523–524, 581
development, 525
fundamental concepts, 516–524
medical mathematics and, 314
potential applications, 528
qualitative, 141, 455, 474
quantitative, 141, 455
real-life applications, 525–528
scientific use of, 474
sports, 380–382, 498
Statues, *432,* 433–434
Stealth technology, 182, 244–245
Steganography, 266–267
Stella II, 330
Stem cell research, 446–447
Stem graphs, 250
Stem and leaf plots, 404, 407
Stepped Reckoner device, 336
Steps, logic, 300–301, 301*t*
Steradian, 123
Stereo sound, 282–283
Stock (Goods), 592
Stocks (Securities)
business math and, 67
distribution of, 150–151
graphing, 259
investments, 191
line charts of, 111
market trends, 515
share values, 436
stock options, 191
zero-sum games and, 597
Stokes, Gabriel, 212
Stonehenge, 97, *233*
Stones (Unit of measure), 129
Stopwatches, 496
Store assistants, 592
Straight bet, 508
Stratified sampling, 462, 517
Strauss, Joseph B., 76
Street signs, 414
Stress, 227
Strictly competitive games, 596
String theory, 555
Strings
guitar, 207, 242
knot theory of, 355
monochord, 344
overtones of, 347
Strõmers, Martin, 126
Structural design, 35–36, 89–90
Student loans, 56–57
Students
grade percentages, 374, 383
grade point average, 139–140, 140*f*
time management for, 153
See also Education
Student-teacher ratio, 445
Submarines, 41, 182, 482
Subnet mask, 118
Sub-populations, 517
Subtraction, 485, **529–536,** *530*
Sumerians, 97, 170
Sunflowers, 354
Super Bowl, 266
Supercomputers, 78
Supersonic flight, 299
Surface area, *47,* 48, 439, 576
Surgery, robotic, 245
Surveying, *49,* 49–50, 514–515
area conversions, 129
quadratic and cubic equations for, 439
trigonometry for, 557, 564–565, *565*
word problems for, 592
Suspension bridges, 76–77
Swedish beds, 42
Swimming pools, 152, 386–387, 446, 581
Symmetric secret key encryption, 120–121, 362
Symmetry, **537–542**
abstract, 542
architectural, 34, *35,* 36–38, *39,* 541, *541*
bilateral, 38
city planning and, 42
fractal, 541–542, 541*f*
glide-reflection, 537
reflection, 38, 537–542, *538*
rotational, 38, 537–542, 539*f,* 540*f*
sliding, 34
textile design and, 43
translational, 34, 38, 537–542, 539, 539*f*
Synthesizers, electronic sound, 213, 347–348

T

Table tops, 37
Tables, **543–552**
  addition, 544, 544*f*
  astronomical, 544
  chord, 561–562
  conversion, 545
  development of, 72, 543–544
  financial, 546–548
  fundamental concepts, 543
  periodic, 545, 546*f*
  potential applications, 549–550
  real-life applications, 544–551
  Rudolphine, 544
  tide, 549
  trigonometry, 544, 561
  types, 544
  See also Multiplication table
Tacoma Narrows bridge, *90*
Taj Mahal, 40
Talmud, Babylonian, 226, 228
Tamoxifen, 325
Tangents, 564
Task-specific rubrics, 455–456
Tasmanian sheep, 331
Tax cuts, 56
Taxes
  apportionment, 209
  geometric representation, 235
  sales, 377–378
  value added, 592
  See also Income taxes
T-distribution, 522
Teacher-student ratios, 445
Technology access, 410, 411*f*
Teenagers, 578
Telephones. *See* Cellular telephones
Telephony, Internet, 119
Telephoto lenses, 399, 402–403
Teleportation, 23–24
Telescopes, 43, 55, 165–166
Television, closed circuit, 31
Temperament
  mean-tone, 346
  well, 346–347
Temperature
  absolute, 127
  absolute zero, 127, 356, 466–467
  average, 152
  Celsius, 123, 125, 340, 357, 466
  conversion of, 125–126, 340
  Fahrenheit, 125–126, 340, 357
  freezing point, 126, 356
  gigaelectron volts, 487
  Kelvin, 123, 125, 126, 127, 357, 467
  negative numbers and, 357
  oven, 127, *127,* 302
  subtraction of, 530
  wind chill and, 476
Temples, 237
Ten-based numbering systems. *See* Decimals
1099 Form, 189
Tennis, virtual, 291
Terminal velocity, 17
Test scores, 524, 589–590
Text code, 116–117
Textile design, 43
Thales, 236
Theoretical probability distribution, 518
Theory of Games and Economic Behavior (Neumann and Morgenstern), 226
Theory of Proportion, 431–432
Thermal energy, 125
Thermodynamics, 125
Thermometers
  electrical, 126
  glass, 126, 580–581
  interval scale, 466
  mercury, 126, 580–581
  oven, 302
Thickness, 162
Third-order equations, 438
Thomson, William, 127
Three Mile Island, 370
365-day calendars, 98, 99
Three-dimensional Cartesian coordinate systems, 133, 133*f,* 136
Three-dimensional images
  animated, 263
  architectural, 513–514
  diagnostic, 475, 480
  graphs, 112, 407
  rotation, 566
  vectors and, 572
Three-dimensional models, 78, 147, 468
Three-dimensional vectors, 569, 569*f,* 572
Three-dimensions, 232, 303
  See also Perspective
Thresholds, 522
Tide tables, 549
Time
  conversion of, 120
  Coordinated Universal Time, 547
  definition, 307
  doubling, 173–174
  elapsed, 3
  estimation of, 161–162, 166
  geologic, 488
  geometry and, 245–247
  Greenwich Mean Time, 135, 547, 547*f*
  measurement of, 311
  navigation and, 135
  rounding off, 452
  scales, 469
  space and, 236, 246–247, 312
  theory of relativity and, 131–132
Time management, 153
Time travel, 246–247
Time zones, 547
Timing systems, sports, 139, 339, 496
Tips, calculating, 194–195, 375–376
Tires, 439–440
Toads of the Short Forest (Zappa), 351
Toll roads, 533
Tomography, computed, 147, 475, 480
Tone
  well-tempered, 346–347
  wolf, 346
Tons, metric, 129
Tools, measuring, 208
Topographic maps, 104, *104,* 136
Topology, **553–556,** *554, 555*
Tornado modeling, 201
Tortoise shells, *354*
Touchdowns, 499, 501–502
Tournaments, 590
Toys, 471
Track and field events, 3, 358, 496, 534
Tracking, human motion, 290–291
Tracks (Race), 309, *312*
Trade-ins, automobile, 187
Translation, 587
Translational symmetry, 34, 38, 537–542, 539*f*
Transposition-substitution algorithms, 29
Travel
  fuzzy logic for, 301–302
  measurement for, 307
  metric system conversion for, 339–340
  time, 246–247, 431
  time estimates, 586
  trip length, 443
  See also Distance
Traveling salesperson, 589
Traverse Mercator projection, *101*
Tree diagrams, 412–413
Trees
  family, 413
  leaves, 200
Trend plots, 259
Trends, 374, 463
Trial division, 150
Trial software, 285
Triangles
  acute, 559
  area of, 340
  description, 233
  equilateral, 237, 541, 558–559, 558*f*
  isosceles, 559
  obtuse, 559
  Pascal's, 500–501, 501*f*
  plane, 558–559
  potential applications, 566

Triangles (*contd.*)
Pythagorean theorem and, 512, 513, *513*
rafters as, 237–238
right, 559–560, 559*f*
scalene, 559, 559*f*
Sierpinski, 541–542, 541*f*
similar, 182
spherical, 559
See also Trigonometry
Triangular graphs, 407, 410–411
Tricycles, 505
Trigonometry, **557–567**
definition, 233
development, 561–562
fundamental concepts, 557–561
Pythagorean theorem and, 474–475, 475*f*
real-life applications, 562–566
tables, 544, 561
three-dimensional imaging and, 263
Trinity (Masaccio), 391
Triple redundancy, 275
Trompe d'oeil, 395–396
Tropical Rainfall Measuring Mission (TRMM), 77
Trucks, 441–442, 451
See also Automobiles; Vehicles
Trump, Donald, 194
Truth and logic, 300, 583
Tsunami, *298,* 471, 591–592
T-test, 522, 525
Turing, Alan, 29
Twain, Mark, 381
Twin prime conjecture, 361
Two by two matrix equations, 288, 288*f*
Two-dimensional vectors, 568–569, 568*f,* 569*f*
Type II diabetes, 315

U

Ultra Deep Field, 55
Unequally likely messages, 271–272
Uniform distribution, 518
Unit fractions, 203–204
United States
Census Bureau, 193, 451
Coast and Geodetic Survey, 564
Congressional Representatives, 209
Constitution, 208–209
Department of Energy, 214, 451
Food and Drug Administration, 245
Geologic Survey, 564
measurement system, 124–125
National Safety Council, 5
National Security Agency, 147
presidential elections, 141–142, 459
See also National Aeronautic and Space Administration
Units of area, 45–46
Universal product code (UPC) barcodes, *15,* 15–16, *23*
Universal Transverse Mercator (UTM), 103
Universe
age of, 160
expanding, 178–179
origin of, 487
time and, 245–247
Unmanned rovers, 388
Unsharp masking, 403
UPC barcodes, *15,* 15–16, *23*
U.S. Coast and Geodetic Survey, 564
U.S. Geologic Survey, 564
Used automobiles, 162–163
Usury rate, 186
UTC (Coordinated Universal Time), 547, 547*f*
UTM (Universal Transverse Mercator), 103

V

Vacancy rates, 384
Vaccines, MMR, 594
Validity, 583
Value added taxes (VAT), 592
Value meals, 4
Vanishing point, 389, 390*f,* 392, 392*f,* 395
Variable expressions, 10
Variables, 211, 249, 328, 329
Variance, 318, 319–320, 520
Variance, analysis of (ANOVA), 522–523
Variogram, 525–526
VAT (Value added taxes), 592
Vectors, 233–234, **568–574,** *570, 573*
algebra for, 570–571, 571*f*
coordinate systems and, 132
matrix equations and, 288, 288*f*
three-dimensional, 569, 569*f,* 572
trigonometry and, 563–564
two-dimensional, 568–569, 568*f,* 569*f*
velocity and, 571–572
Vehicles
aerodynamic and hydrodynamics of, 259
crash tests of, 18–19
robotic, 388
weight of, 535
See also Automobiles
Velocity
integral of, 83–84, 84*f*
measurement, 307
of pitches, 139
spacetime and, 136
terminal, 17
trigonometry and, 563–564
vectors, 571–572, *573*
See also Speed
Venn diagrams, 493, 493*f*
Verhulst, P. E., 200
Vermerer, Johannes, 391
Vertex, 590, 590*f*
Vertical jumps, 534
Vibration, 479
Video analysis
human motion, 290–291
sports, 263
Video file size, 118
Video morphing, 134
Viete, François, 13, 139, 279, 417
Viking 1, 265
Viking 2, 265
VIOP (Voice over Internet protocol), 155
Viral marketing, 341
Virtual-reality, 291, 412
Viruses, 355
Vision, 281
Visual geometry, 35–36
Visualization software, 258, 260, 260*f*
Vitruvian Man (da Vinci), 432, *433*
Voice over Internet protocol (VIOP), 155
Volume, **575–582,** *577*
body, 439
of boxes, 575
brain, 578, 581–582
calculation of, 124, 129, 575–576, 576*f*
conversion of, 340
of cubes, 439, 575
of the Earth, 162
human body, 439
liquid, 340
quadratic and cubic equations for, 439
of solid objects, 417
Volunteer sampling, 458
Von Leibniz, Wilhelm. *See* Leibniz, Gottfried Wilhelm von
Voting
Electoral College, 209
fractional, 209
patterns, 383
polls, 379, 459, 527–528
preferences, 142
See also Elections
Voyager spacecraft, 276

W

W-2 Form, 189
Wagering. *See* Gambling
Wages. *See* Income; Pay
Waitstaff tips, 194–195, 375–376
Wallace, Alfred Russel, 174–175
Walt Disney Company, 393
War, 597
War (Card game), 595
Warranty period, 427–428
Washington Monument, 238, *239*

Water
freezing point, 126
oil on, 162
quantity of, 473
safety of, 446
wells, 594
Water bodies, perimeter of, 386–387
Water pollution, 462
Watermarks, digital, 266–267
Waves
seismic sea, *298,* 471, 591–592
sound, 494
Tsunami, *298,* 471, 591–592
Weapon guidance systems, 440
Weather
atmospheric pressure, 469
balloons, 477–478
calculators for, 77–78
climate change, 48–49, 257–258, 580, 594
data collection for, 477–478
finite-element modeling of, 212–213
fractals for, 355
maps, 414
modeling, 201
percentages and, 374
relationships and patterns, 474, 476–477
sampling, 461
temperature conversion for, 126–127, 340
wind chill, 476
word problems for, 593
Web crawlers, 146
Weibull, Waloddi, 521
Weibull formula, 521
Weierstrauss, Karl, 156
Weight
friction and, 563–564
height and weight charts, 319–320, 548–549, 548*f,* 549*f*
measurement of, 129, 308, 340
optimal, 446
rounding off, 450
scales for, 468–469
to-height ratio, 446
of trucks, 451
of vehicles, 535
Weight loss, 534
Weighted average, 54–55, 153
The Well Tempered Clavier (Bach), 347
Wells, water, 594
Well-tempered tone, 346–347
Western music, 351
Wheat, 474
Wheels, 43
WHO (World Health Organization), 591
Whole numbers, 449–450
Wide angle lenses, 399
Wildfire models, 75
Wildlife
body size of, 576
counting, 147–148
endangered, 587
game theory and, 229–230
legs, 179
photographs, 402–403
population dynamics, 305
population sampling, 164
See also Population growth
Wiles, Andrew, 362
Wind chill, 476
Wind strength, 469
Wind tunnels, 259
Windows, 42, 233–234
Wireless service. *See* Cellular telephones
Wolf-tones, 346
Wood chucks, 337
Wooded ecosystems, 201
Woods, Tiger, 285
Word problems, **583–594**
development of, 584
fundamental concepts, 583–584
real-life applications, 584–594
World Health Organization (WHO), 591
World Series, 427
World War II military model, 332–335, 333*f*
World Wide Web (WWW), 146, 412
See also Internet
Writing, secret. *See* Cryptography

X

X axis, 107
XOR algorithm, 362
XP (Extreme programming), 285
X-rays, 147, 264, 414
x-y graphs, 254–257, 254*f*
x-y scatter graphs, 109, 109*f*

Y

Y axis, 107
Yard (Measurement), 308, 340, 579
Yards gained (Football), 499
Years
leap, 99
light, 486–487
Yen, 196
Yuan, 196
Yucca Mountain, 175–177, *176,* 213–214

Z

Zadeh, Lotfi, 302
Zappa, Frank, 351
Zeno of Elea, 492
Zero
absolute, 127, 356, 466–467
division by, 150
negative numbers and, 356
silence and, 345–346
Zero-sum games, 226, **595–597,** *596*
Zeta function, 212
Zipped files, 119